

QY 353 PROXGHIISGLUASPMETThrleu 361  
 DB 664 CCCCAGCAGAGACATGACCTGCTA 690

RESULT 15  
 BG715206  
 LOCUS 602675668f1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4798145,  
 DEFINITION mRNA sequence.  
 ACCESSION BG715206 702 bp mRNA linear EST\_08-MAY-2001  
 VERSION BG715206  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 702)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@dbp-remail.nih.gov  
 Tissue procurement: Miklos Palkevics, M.D., Ph.D.  
 cDNA library preparation: Michael J. Brownstein (NHGRI), Shireki  
 Toshiyuki and Piero Carninci (RIKEN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLM10685 row: h column: 18  
 High quality sequence stop: 702.

FEATURES  
 source location/Qualifiers  
 1..702  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4798145"  
 /clone\_lib="NIH MGC 96"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pBluescript (modified  
 pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
 ); Oligo-dT primed using primer 5'-TTTTTTTCTTTTNN-3',  
 size-selected for average insert size 2.3 kb and  
 normalized to 10^5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 151 a 195 c 220 g 136 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2,93e-105 Length: 702  
 Score: 1129.00 Matches: 224  
 Percent Similarity: 99.12% Conservative: 0  
 Best Local Similarity: 99.12% Mismatches: 1  
 Query Match: 43.76% Indels: 1  
 DB: 12 Gaps: 0

US-09-830-144-4 (1-504) x BG715206 (1-702)

QY 1 MetAlaAlAGlnArgSerleuInserglugInlProserTrpThraasp 20  
 DB 26 ATGGCGCGGAGGAGGAGCTTGCTGCAGAGTGAGCGAGCCAACTGACAGATGAC 85

QY 21 LeuProLeuCyHisleuSerGlyValGlySerAlaSerAsnArgSerTrpSerAlaasp 40  
 DB 86 CTGCTCTCTGCGACCTATCTGAGGATGAGCTCAGCTCCAGCCGACGCTACTGCTGAT 145

QY 41 GlyLysGlyThrGlnSerHisProGluAspSerTrpLeuLysPheArgSerGluasn 60

DB 146 GGCAGGGCACTGAGAGCCACCCGCCAGAGAGACAGCTGCTCAAGTTCAGAGTACAGAC 205  
 QY 61 AasnCySphenleuTrpGlyValPheasnGlyTrpAspGlyAsnArgValThrAsnheVal 80  
 DB 206 AACTCTCTCTGATGAGGAGGCTTCTCAACGAGCTATGATGAGCAACGAGTACCAACTTCCTG 265

QY 81 AlaGlnArgLeuSerAlaGlnLeuLeuGlyGlnLeuAsnAlaGlnHisAlaGlnAla 100  
 DB 266 GCCCGAGGCTGCTGCGAGAGCTTCTGAGGCTGAGCTGAAATGCCAGACACCCGAGGCC 325

QY 101 AspValArgArgValLeuLeuGlnAlaPheAspValGluArgSerPheLeuGlnSer 120  
 DB 326 GATGTGCGGCGCTGCTGCTGCGAGAGCTTCTGATGTGTGAGAGAGCTTCTGAGAGTCC 385

QY 121 IleAspAspAlaLeuAlaGlnLysAlaSerLeuGlnSerGlnLeuProGlnGlyValPro 140  
 DB 386 ATTGACGAGCGCTTGCTGCTGAGAAAGCAAGCTTCAATGCCAAATGCCAGAGGAGTCCCT 445

QY 141 GlnHisGlnLeuProProGlnTrpGlnLysIleLeuGlnArgLeuLysThrLeu-GluArg 160  
 DB 446 CAGCACCGAGCTGCTGCTCTCTCAATCAAGATCTTGAAGAGACTCAAGACGTTCAAGAGAG 505

QY 160 GGUlleserGlyGlyAlaMetAlaValAlaValLeuLeuAsnAsnLysLeuTrpVal 180  
 DB 506 GGAATTTCCGAGAGGAGGCGCATGGCGTTGTGGCGTCTTCTCAACAAAGACTTACT 565

QY 180 AlaAsnValGlyThrAsnArgAlaLeuLeuCySlysSerThrValAspGlyLeuGlnVal 200  
 DB 566 CCCCAATGCTGCTGCTCAACCGCTGCACTTTATGCAATGAGACAGTGGATGGCTTGCAGGT 625

QY 200 ThrGlnLeuAsnValAspHisThrThrGlnAsnGlnAspGlnLeuPheArgLeuSerGln 220  
 DB 626 GACACAGCTGAACGTCAGACCAACCAACAGAGATGAGCTTCTTCCGCTTTCGCA 685

QY 220 nLeuGlyLeuAspAla 225  
 DB 686 GCTGGAGCTTGATGCT 701

Search completed: December 10, 2002, 02:04:05  
 Job time : 1587.89 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 9, 2002, 23:00:44 (Search time 53.518 Seconds  
(without alignments)  
2888.095 Million cell updates/sec

Title: US-09-830-144-4  
Perfect score: 2580  
Sequence: 1 MAORRSLIQEQPSWTD.....AEFYRLMSVDHGQSVTAP 504

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODBL=frame+ p2n model -DEV=xld  
-Q=/cgn2\_1/USPTO\_spool/US09830144/rumat\_04122002\_141354\_2283/app\_query\_fasta\_1.1422  
-DB=Issued\_Patents\_NA -OPMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPTCL=0  
-LOOPTXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdt  
-LIST=45 -DOCALLIGN=200 -NORM=SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09830144@cgn2\_1.144 @rnat\_04122002\_141354\_2283 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	2580	100.0	1560	US-08-752-891-1	Sequence 1, Appl1
2	2580	100.0	1560	US-09-144-178-1	Sequence 1, Appl1
3	2580	100.0	1560	US-09-406-854-1	Sequence 1, Appl1
4	2580	100.0	1560	US-09-529-279-1	Sequence 1, Appl1
5	2580	100.0	1560	US-09-529-279-10	Sequence 1, Appl1
6	2577	99.9	1560	US-09-529-279-42	Sequence 42, Appl1
7	2575	99.8	1560	US-08-752-891-5	Sequence 5, Appl1
8	2575	99.8	1560	US-09-144-178-5	Sequence 5, Appl1
9	2575	99.8	1560	US-09-406-854-5	Sequence 5, Appl1
10	178	6.9	1403	US-09-013-881-10	Sequence 10, Appl1
11	131	5.1	49377	US-08-764-233A-1	Sequence 1, Appl1
12	128.5	5.0	1824	US-08-822-701-1	Sequence 1, Appl1

13	128.5	5.0	1824	3	US-08-935-855-1	Sequence 1, Appl1
14	128.5	5.0	2081	3	US-08-935-855-21	Sequence 21, Appl1
15	126.5	4.9	28958	1	US-08-258-261B-6	Sequence 6, Appl1
16	126.5	4.9	28958	1	US-08-456-837-6	Sequence 6, Appl1
17	126.5	4.9	28958	1	US-08-457-342-6	Sequence 6, Appl1
18	126.5	4.9	28958	1	US-08-457-646A-6	Sequence 6, Appl1
19	126.5	4.9	28958	1	US-08-457-646A-6	Sequence 6, Appl1
20	126.5	4.9	28958	1	US-08-457-646A-6	Sequence 6, Appl1
21	126.5	4.9	28958	1	US-08-764-233A-4	Sequence 4, Appl1
22	126.5	4.9	28958	1	US-08-457-335A-6	Sequence 6, Appl1
23	126.5	4.9	28958	1	US-08-729-214-6	Sequence 6, Appl1
24	125.5	4.9	28958	3	US-08-935-855-19	Sequence 19, Appl1
25	121	4.7	68750	3	US-09-335-409-1	Sequence 1, Appl1
26	121	4.7	68750	4	US-09-568-102-1	Sequence 1, Appl1
27	121	4.7	68750	4	US-09-567-969-1	Sequence 1, Appl1
28	121	4.7	68750	4	US-09-568-480-1	Sequence 1, Appl1
29	121	4.7	68750	4	US-09-568-480-1	Sequence 1, Appl1
30	121	4.7	68750	4	US-09-568-477-1	Sequence 1, Appl1
31	121	4.7	68750	4	US-09-567-899-1	Sequence 1, Appl1
32	117	4.5	3182	1	US-08-188-582-12	Sequence 12, Appl1
33	117	4.5	3182	1	US-08-646-715-12	Sequence 12, Appl1
34	117	4.5	71989	4	US-09-443-501A-2	Sequence 2, Appl1
35	116.5	4.5	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
36	116.5	4.5	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
37	116	4.5	1785	2	US-08-910-856-1	Sequence 1, Appl1
38	116	4.5	3187	2	US-08-910-856-9	Sequence 9, Appl1
39	116	4.5	3187	2	US-08-910-856-10	Sequence 10, Appl1
40	115.5	4.5	4660	4	US-09-166-350-26	Sequence 26, Appl1
41	115	4.5	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
42	115	4.5	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
43	113.5	4.4	44377	2	US-08-804-227C-7	Sequence 7, Appl1
44	113.5	4.4	44377	2	US-08-804-199-1	Sequence 1, Appl1
45	111.5	4.3	11601	2	US-08-222-617A-3	Sequence 3, Appl1

## ALIGNMENTS

RESULT 1  
US-08-752-891-1  
Sequence 1, Application US/08752891  
Patent No. 5837819  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihiko  
TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESS: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,891  
FILING DATE: 20-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300  
 ; TELEFAX: (202)672-5399  
 ; TELEX: 904136  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1560 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 30..1541  
 ; FEATURE: mat\_peptide  
 ; NAME/KEY: mat\_peptide  
 ; LOCATION: 30..1541  
 ; US-08-752-891-1

Alignment Scores:  
 Pred. No.: 4,82e-249 Length: 1560  
 Score: 2580.00 Matches: 504  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 2 Gaps: 0

US-09-830-144-4 (1-504) x US-08-752-891-1 (1-1560)

QY 1 MetAlaAlaGlnArgSerLeuLeuGlnSerGluGlnProSerTrpThrAspAsp 20  
 DB 30 ATGGCGGCGCAGAGAGAGAGCTTGCAGAGTGAGCAGCAGCCCAAGCTGCACAGATGAC 89  
 QY 21 LeuProLeuGlnHisLeuSerGlyValGlySerAlaSerAsnArgSerTrpSerAlaAsp 40  
 DB 90 CTGCCCTCTCTCCCACTCTCTGGGGTGGCTCAGCCTCCCAACCGCAGACTCTCTGCTGAT 149  
 QY 41 GlyLysGlyThrGluSerHisProGluAspSerTrpLeuLysPheArgSerGluAsn 60  
 DB 150 GGCAGAGGCACTGAGAGCCACCCGCCAGAGGACAGCTGGCTCAAGTTCAGGAGTGAGAAC 209  
 QY 61 AsnGlnPheLeuGlyValPheAsnGlyTrpAspGlyAsnArgValThrAsnPheVal 80  
 DB 210 AACTGCTTCTCTGATGGGGTCTTCAACGGCTATGATGGCAACCGAGTGACCACTTCGTG 269  
 QY 81 AlaGlnArgLeuSerAlaGluLeuLeuLeuGlnLeuAsnAlaGluHisAlaGluAla 100  
 DB 270 GCCAGACGGCTGTCGACAGAGCTCTGCTGGCCAGCTGAATGCCAGCAGCAGCCGAGGCC 329  
 QY 101 AspValArgArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGluSer 120  
 DB 330 GATGTGGGCGGTGTCGTCTGAGGCTTCGATGTGGTGAGAGAGGCTTCTTGAGTGCC 389  
 QY 121 IleAspAspAlaLeuAlaGluLysAlaSerLeuGlnSerGlnLeuProGluGlyValPro 140  
 DB 390 ATTGACGACGCTTGGCTGGAAGGCAAGCCTCCAGTCCCAATTGCCAGAGGAGTCCCT 449  
 QY 141 GlnHisGlnLeuProGlnTrpGlnLysIleLeuGluArgLeuLysThrLeuGluArg 160  
 DB 450 CAGCACCACTGCTCTCTAGTATCAGAGATCTTTCAGAGACTCAAGACCTTAGAGAGG 509  
 QY 161 GluLysGlyGlyAlaMetAlaValAlaValLeuLeuAsnAsnLysLeuTrpVal 180  
 DB 510 GAAATTTGGGAGGGGCGCATGGCGGTGTGGCGGTCTTCTCAACAACAAGCTTACGTC 569  
 QY 181 AlaAsnValGlyThrAsnArgAlaLeuLeuGlnLysSerThrValAspGlyLeuGlnVal 200  
 DB 570 GCCAATGTCGTAACAACCGTGACCTTTATGCAATCGACAGTGGATGGGTTCAGGTG 629  
 QY 201 ThrGlnLeuAsnValAspHisThrThrGluAsnGluAspGluLeuPheArgLeuSerGln 220  
 DB 630 ACACAGCTGAACGTGGACCAACACACAGAGAGAGAGTCTTCCGTCTTCGACG 689  
 QY 221 LeuGlyLeuAspAlaGlyLysIleLysGlnValGlyIleLysCysGlyGlnGluSerThr 240

DB 690 CTGGCTTGGATGCTCGAAGATCAAGCAGGTGGGATCATCTGTGGCAGGAGACACC 749  
 QY 241 ArgArgIleGlyAspTrpLysValLysTrpGlyTrpThrAspIleAspLeuLeuSerAla 260  
 DB 750 CGCGGATCGGGATTACAAAGTTAAATATGGCTACAGGACATTGACCTTCTCAGCGCT 809  
 QY 261 AlalysSerLysProIleAlaGluProGluIleHisGlyAlaGlnProLeuAspGly 280  
 DB 810 GCCAAGTCCAACCAATCATCGACAGCCAGAAATCCATGGGGCACAGCCGCTGGATGGG 869  
 QY 281 ValThrGlyPheLeuValLeuMetSerGluGlyLeuTrpLysAlaLeuGluAlaHis 300  
 DB 870 GTGACGGGCTTCTGGTGTGATGTCGGAGGGTGTACAGAGCCCTAGAGGAGCCCAT 929  
 QY 301 GlyProGlyGlnAlaAsnGlnGluAlaMetIleAspThrGluPheAlaLysGln 320  
 DB 930 GGGCTGGGCGAGCCCAACCGAGATTTGTCGATGATTGACACTGAGTTTGGCAAGCAG 989  
 QY 321 ThrSerLeuAspAlaValAlaGlnAlaValValAspArgValLysArgIleHisSerAsp 340  
 DB 990 ACCTCCTTGGACGAGTGGCCCGAGCCCTCTGGACCGGGTGAAGCGCATCCACAGCGAC 1049  
 QY 341 ThrPheAlaSerGlyGlyGluArgAlaArgPheCysProArgHisGluAspMetThrLeu 360  
 DB 1050 ACCTTCGCCAGTGTGGGAGCGTCCAGGTTCTGCCCGGCGACGAGCATGACCCCTG 1109  
 QY 361 LeuValArgAsnPheGlyTrpProLeuGlyGluMetSerGlnProThrProSerProAla 380  
 DB 1110 CTAGTGAGGAACCTTTGGCTACTCCCGCTGGCGAATGAGCCAGCCACACCGAGCCGACC 1169  
 QY 381 ProAlaAlaGlyGlyArgValTrpValSerValProValSerValProValSerVal 400  
 DB 1170 CCAGCTGCAGGAGGAGTGTACCTGTGTGTGCTACTTCCAGCGCCGAGCAGCAGC 1229  
 QY 401 SerLysThrSerValThrLeuSerLeuValMetProSerGlnGlyGlnMetValAsnGly 420  
 DB 1230 AGCAAGACCAAGCTGACCTCTCTTGTGTGCTCCCGAGGCGCCAGATGGTCAACGGG 1289  
 QY 421 AlaHisSerAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThr 1440  
 DB 1290 GCTCACAGTCTTCCACCTTGAGGAGCCACCCCTTCCACCAACCAAGCCGAGCC 1349  
 QY 441 LeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerSerSerGly 460  
 DB 1350 TTAACCTGTCAGTCCACCAACACGACACGACGACGACGACGACGACGACGACGACG 1409  
 QY 461 LeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGlyArgValGluPro 480  
 DB 1410 CTCTTCCGCTCCCGGCCGCCACTCGCTCCGCTCCGCGGAGGAGCGTCTGTGTGAGCCC 1469  
 QY 481 TyrValAspPheAlaGluPheTrpArgLeuTrpSerValAspHisGlyGlnSerVal 500  
 DB 1470 TATGTGACTTTGCTGAGTTTACCGCTCTGGAGCGTGGACCATGGCGAGCAGAGCGTG 1529  
 QY 501 ValThrAlaPro 504  
 DB 1530 GTGACAGCACCG 1541

# RESULT 2

US-09-144-178-1  
 ; Sequence 1, Application US/09144178  
 ; Patent No. 5989862  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MATSUMOTO, Kunihiro  
 ; APPLICANT: NISHIDA, Eisuke  
 ; TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA



```

ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/144,178
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/752,891
FILING DATE: 20-NOV-1996
APPLICATION NUMBER: JP 8-300856
FILING DATE: 28-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-126282
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17981/111
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SRO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1560 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1541
FEATURE:
NAME/KEY: mat peptide
LOCATION: 30..1541
US-09-144-178-1
Alignment Scores:
Pred. No.: 4.82e-249 Length: 1560
Score: 2580.00 Matches: 504
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 2
US-09-830-144-4 (1-504) x US-09-144-178-1 (1-1560)
QY 1 MetAlaAlaGlnAArgSerLeuGlnSerGluGlnGlnProSerThrPThrAspAsp 20
Db 30 ATGGGGGGGAGAGAGAGAGCTTGCTGCAGAGTGCAGCGCAGCCAGCTGAGCAGATGAC 89
QY 21 LeuProLeuCySHisLeuSerGlyValGlySerAlaSerAlaSerAlaSerAlaSerAla 40
Db 90 CTGCTCTCTGCTGCACTCTCTGCGGTTGGCTCAAGCTCCAAACCCAGTATCTGCTGAT 149
QY 41 GlyGlyGlyThrGluSerHisProProGluAspSerThrLeuLeuPheArgSerGluAsn 60
Db 150 GGCAGAGGCGACAGAGAGCCAGCCGCGCAGAGAGCAGCTGGCTCAAGTTCCAGAGTGAAC 209
QY 61 AsnGlyPheLeuThrGlyValPheAsnGlyThrAspGlyAsnAsnValThrAsnPheVal 80
Db 210 AACTGCTTCTGTATGGAGGCTTTCACGAGCTATGATGCAACCGAGTGACCAACTTGTG 269
QY 81 AlaGlnArgLeuSerAlaGluLeuLeuGlnLeuAsnAlaGluHisAlaGluAla 100
Db 270 GCCCAGCGGCTCTCCGAGAGAGCTCTGCTGGCCAGCTGAGTGCAGACAGCCGAGGCC 329
QY 101 AspValArgArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGluSer 120
Db 330 GATGTGCGGCGTGTGCTGCTGCAGAGCCTTCGATGTGTGTGAGAGAGAGCTTCTGTGAGTCC 389

```

```

QY 121 IleAspAspAlaLeuAlaGluLeuAlaSerLeuGlnSerGlnLeuProGluGlyValPro 140
Db 390 ATTGACGAGCGCTTGCTGAGAGAGCAAGCTCCAGTGCAGATTTGCCAGAGAGTCCCT 449
QY 141 GlnHisGlnLeuProProGlnThrGlnLysIleLeuGluArgLeuLysThrLeuGluArg 160
Db 450 CAGCAGCCAGCTGCTCTCCAGTATCAGAAAGATCCTTGAGAACTCAAGACCTTGAAGAG 509
QY 161 GluIleSerGlyGlyAlaMetAlaValAlaValLeuLeuAsnAsnLeuThrVal 180
Db 510 GAAATTTGGAGGGGGCCATGAGCGCTGTGGGCTTCTTCAACACCAAGCTTACGTC 569
QY 181 AlaAsnValGlyThrAsnArgAlaLeuLeuCySHisSerThrValAspGlyLeuGlnVal 200
Db 570 GCCATGTGCGACAAACCGTGCACTTTATGCAAAATGACAGTGGATGGGTTCAGGTG 629
QY 201 ThrGlnLeuAsnValAspHisThrThrGluAsnGluAspGluLeuPheArgLeuSerGln 220
Db 630 ACACAGCTGAACGTGACACACACACAGAGAACGAGATGAGCTTCTCCGTTTCGAG 689
QY 221 LeuGlyLeuAspAlaGlyLysIleLysGlnValGlyIleLysGlyGlnGluSerThr 240
Db 690 CTGGGCTTGATGCTGGAAGATCAAGAGGTGGGATCATCTGTGSCAGAGAGCACCC 749
QY 241 ArgArgIleGlyAspThrLysValLysThrGlyThrThrAspIleAspLeuLeuSerAla 260
Db 750 CGCGGATGGGGATTACAGGTTAAATATGCTTACAGGACATTGACCTTCTCAGCGCT 809
QY 261 AlaLysSerLysProIleLeuAlaGluProGluLysIleGlyAlaGlnProLeuAspGly 280
Db 810 GCCAAGTCCAAACCAATCATCCAGAGCCAGAAATCCATGGGGGCACCCCTGTGATGG 869
QY 281 ValThrGlyPheLeuValLeuMetSerGluGlyLeuThrLysAlaLeuGlnAlaHis 300
Db 870 GTGACGGGCTTCTGTGCTGATGTGCGAGGGGTTGTCAAGCCCTTGAAGAGCCCAT 929
QY 301 GlyProGlyGlnAlaAsnGlnLysIleAlaMetIleAspThrGluPheAlaLysGln 320
Db 930 GGGCTGGGCAAGCCAGCAGAGAGATGCTGCGATGATTCACCTGATTTGCCAAGCAG 989
QY 321 ThrSerLeuAspAlaValAlaGlnAlaValAlaAspArgValLysArgIleHisSerAsp 340
Db 990 ACCTCCCTGAGCAGAGTGGCCAGGCCGCTGCGAGCCGGGTGAGCCGATCCACAGCCAC 1049
QY 341 ThrPheAlaSerGlyGlyGluArgAlaArgPheCyProArgHisGluAspMetThrLeu 360
Db 1050 ACCTTGCCAGTGTGGGAGCGTGCAGGTTCTGCCCGGACAGAGAGCATGACCTTG 1109
QY 361 LeuValArgAsnPheGlyThrProLeuGlyGluMetSerGlnProThrProSerProAla 380
Db 1110 CTAGTAGAAGAACTTGTGCTACCCGCTGGCGAATGAGCCAGCCACAGCCAGCC 1169
QY 381 ProAlaIleGlyValArgValLysProValSerValProThrSerSerAlaGlnSerThr 400
Db 1170 CCAGCTGCAGAGAGAGAGATGTATCCCTGTGTGCTGCTATCCAGGCGCCAGAGCAC 1229
QY 401 SerLysThrSerValThrLeuSerLeuValMetProSerGlnGlyGluMetValAsnGly 420
Db 1230 AGCAAGACCAAGCTGACCTCTCTGCTGATGCTCCACAGGCGCAGATGATCAAGGG 1289
QY 421 AlaHisSerAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThr 440
Db 1290 GCTCAGAGTCTTCCACCTTGACGAGAGCACCCCACTTCCACCAACCAAGCCGAGCC 1349
QY 441 LeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerAspGlyGly 460
Db 1350 TTAACCTCGAGTCCACCAACAGCAGCAGAGAGAGAGCTTCCAGCTTACCGAGAGCC 1409
QY 461 LeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGlyValArgValGluPro 480
Db 1410 CTCTTCGCTCCCGGCCGCGCCACTGCTCCGCTTGGCGAGAGAGCGGTGTTAGGCC 1469

```



Db 1110 CTAGTGGAGAACTTTGGCTACCGCTGGGGAAATGACCAAGCCACACCGACCCAGCC 1169  
Qy 381 ProAlaAlaGlyArgValTyrProValSerValProTyrSerSerAlaGlnSerThr 400  
Db 1170 CCAAGCTCAGAGAGACAGATGTACCTGTGTCTGTGCATATCTCAGCGCCAGAGACCC 1229  
Qy 401 SerTyrSerSerValThrLeuSerLeuValMetProSerGlnGlyMetValAsnGly 420  
Db 1230 AGCAAGACCAAGCGACCTCTCTCCCTGTGATGCTCCCAAGGGCCAGATGTCAAGGG 1289  
Qy 421 AlaHisSerAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThr 440  
Db 1290 GCTACAGAGCTTCCACCTCGAGAGAACCAACCCCACTCAGCAACCAAGCCGACCC 1349  
Qy 441 LeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerAspGlyGly 460  
Db 1350 TTAAACCTGCAGCTCCACCAACACGACAGCAGCAGCTCAGCTCAGCGAGGC 1409  
Qy 461 LeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGlyArgValGluPro 480  
Db 1410 CTCCTCCGCTCCGAGCCCGCCACTGCTCCGCTGAGGAGCGGTGCTTGAGGCC 1469  
Qy 481 TyrValAspPheAlaGluPheTyrArgLeuTyrSerValAspHisGlyGlnGlnSerVal 500  
Db 1470 TATGTGACTTGTGCTGAGTTTACCGCTCTGAGCGTGGACCATGGCGAGCAGAGCGTG 1529  
Qy 501 ValThrAlaPro 504  
Db 1530 GTGACAGCACCG 1541  
RESULT 4  
US-09-529-279-1  
Sequence 1, Application US/09529279  
Patent No. 6451617  
GENERAL INFORMATION:  
APPLICANT: ONO, KOICHIRO  
APPLICANT: OHTOMO, TOSHITAKO  
LEUCANT: TSUCHIYA, MASAYUKI  
TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES  
FILE REFERENCE: 053466/0278  
CURRENT APPLICATION NUMBER: US/09/529,279  
CURRENT FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: PCT/JP98/04796  
PRIOR FILING DATE: 1998-10-22  
PRIOR APPLICATION NUMBER: JP 9/290188  
PRIOR FILING DATE: 1997-10-22  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 1560  
TYPE: DNA  
ORGANISM: Homo-sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (30) .. (1541)  
US-09-529-279-1  
Alignment Scores:  
Pred. No.: 4.82e-249 Length: 1560  
Score: 2580.00 Matches: 504  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
US-09-830-144-4 (1-504) x US-09-529-279-1 (1-1560)

Db 90 CTGCTCTCTGCACTCTCTGGGGTTGGCTCAGCTCCAAACCCAGCTACTGCTGAT 149  
Qy 41 GlyLysGlyThrGlnSerHisProProGluAspSerThrLeuLysPheArgSerGluAsn 60  
Db 150 GGCAAGGACACAGAGACACCCCGCAGAGACAGCTGGCTCAAGTTCAAGAGTGAAGAC 209  
Qy 61 AsnGlyPheLeuTyrGlyValPheAsnGlyTyrAspGlyAsnArgValThrAsnPheVal 80  
Db 210 AACTGCTTCTTATAGGGGTCTTCAACGGCTATGATGGCAACCAAGTACCAACTCTCTG 269  
Qy 81 AlaGlnArgLeuSerAlaGluLeuLeuGlnGlnLeuAsnAlaGlnHisAlaGlnAla 100  
Db 270 GCCCAGCGGCTCTCCGACAGAGCTCTGCTGGGCCAGCTGAATGCGAGACGCGAGGCC 329  
Qy 101 AspValArgArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGlnSer 120  
Db 330 GATGTGCGCGCTGTGTCTGTGTGAGCGCTTCATGTGTGGAGAGAGCTTCTTGAGATCC 389  
Qy 121 IleAspAspAlaLeuAlaGluLysAlaSerLeuGlnSerGlnLeuProGlyValPro 140  
Db 390 ATTGACAGCGCTTGTGCTGAGAGGAAAGCTTCCAGTGCATTTGCCAGAGGACTCCCT 449  
Qy 141 GlnHisGlnLeuProProGlnTyrGlnLysIleLeuGlnArgLeuLysThrLeuGlnArg 160  
Db 450 CAGCACACACTGCTCTCCCTCAGTATCAGAAAGATCTTGAGAGACTCAAGACGTTAGAGAG 509  
Qy 161 GluLysSerGlyAlaMetAlaValAlaValLeuLeuAsnAsnLysLeuTyrVal 180  
Db 510 GAAATTTCCGGAGGGGCCATGCGCTTGTGGCGGTCTTCCAAACAAAGCTCTACACTC 569  
Qy 181 AlaAsnValGlyThrAsnArgAlaLeuLeuGlnLysSerThrValAspGlyLeuGlnVal 200  
Db 570 GCCAATGTCGTACAAACCTGACCTTTATCAATGACAGTGAAGGTTGAGGTG 629  
Qy 201 ThrGlnLeuAsnValAspHisThrThrGlnAsnGluAspGluLeuPheArgLeuSerGln 220  
Db 630 ACACAGCTGAACGTGGACCAACACACAGAGAGAGATGAGACTTCTCCGTCTTCCGAG 689  
Qy 221 LeuGlyLeuAspAlaGlyLysIleLysGlnValGlyIleIleCysGlyGlnGlnSerThr 240  
Db 690 CTGGCTTGATGCTGGAAGAAGATCAAGCAGTGGGATCATCTGTGGGAGAGAGACCC 749  
Qy 241 ArgArgIleGlyAspTyrLysValLysTyrGlyTyrThrAspIleAspLeuSerAla 260  
Db 750 CGCGGATCGGGATTACAGGTAAATGTGCTACACGACATTAACCTTCTCAGCGCT 809  
Qy 261 AlaLysSerLysProIleIleAlaGluProGluIleHisGlyAlaGlnProLeuAspGly 280  
Db 810 GCCAAGTCCAAACCATCATCCGACAGCCAGAAATCCATGGGGCACAGCGCTGATGG 869  
Qy 281 ValThrGlyPheLeuValLeuMetSerGlnGlyLeuTyrLysAlaLeuGlnAlaHis 300  
Db 870 GTGACGGGCTTCTTGTGTGATGTGAGGGGTTGTACAAAGCCCTAGAGACACCCAT 929  
Qy 301 GlyProGlyGlnAlaAsnGlnGlnIleAlaMetIleAspThrGlnPheAlaLysGln 320  
Db 930 GGGCTTGGGACGACCAACAGAGATGCTGCATGATTAAGACATGAGTTTCCAAAGCAG 989  
Qy 321 ThrSerLeuAspAlaValAlaGlnAlaValAlaAspArgValLysArgIleHisSerAsp 340  
Db 990 ACCCTCCCTGACCGCAGTGGCCAGGCCGCTGTGAGACCGGCTGAAGGCATCCACAGCGAC 1049  
Qy 341 ThrPheAlaSerGlyGlyGluArgAlaArgPheCysProArgHisGlnAspMetThrLeu 360  
Db 1050 ACCCTGCGCAGGTGTGGGGAGCGTCCAGATTCTGCCCCGCGACAGAGACATGACCTGT 1109  
Qy 361 LeuValArgAsnPheGlyTyrProLeuGlyGluMetSerGlnProThrProSerProAla 380  
Db 1110 CTAGTGGAGAACTTTGGCTAACCGCTGGGCGAAATGAGCAGGCCACACCGGACCCAGCC 1169  
Qy 381 ProAlaAlaGlyArgValTyrProValSerValProTyrSerSerAlaGlnSerThr 400

Db	1170	CCAGCTCCAGGAGCAGAGTGTTACCTGTGTCTGTGCCATCTACTCCAGCGCCCCAGAGCACC	1229
Qy	401	SerLysThrSerValThrLeuSerLeuValMetProSerGlnGlyGlnMetValAsnGly	420
Db	1230	AGCAAGACACAGCTGACCTCTCCCTTGTTCATGCCCTCCAGGGCCAGATGTTCAACGGG	1289
Qy	421	AlaHisSerAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThr	440
Db	1290	GCTCACAGTCTCTCCACCTTGGACGAAGCACCCTCCACCTCACCACCAAAGCCCGACC	1349
Qy	441	LeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerAspGlyGly	460
Db	1350	TTAACCTTCGAGTCCACCAACACGACACGCGAGACGACGCTCCAGCTCTTCAGCGAGGC	1409
Qy	461	LeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGlyArgValGluPro	480
Db	1410	CTCTTCGGCTCCCGGCCGCCCACTCGCTCCCGCTCGCGAGGACGGTGTGTGTAGCCCC	1469
Qy	481	TyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGlyGluGlnSerVal	500
Db	1470	TATGTGGACTTGTCTGAGTTTTACCGCTCTCGAGCGTGGACCATGGCGAGCAGAGCGTG	1529
Qy	501	ValThrAlaPro	504
Db	1530	GTGACAGCACCG	1541

## RESULT 5

```

US-09-529-279-10
; Sequence 10, Application US/09529279
; Patent No. 6451617
; GENERAL INFORMATION:
; APPLICANT: ONO, KOICHIRO
; APPLICANT: OHTOMO, TOSHIHIKO
; APPLICANT: TSUCHIYA, MASAYUKI
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES
; FILE REFERENCE: 053466/0278
; CURRENT APPLICATION NUMBER: US/09/529,279
; PRIOR FILING DATE: 2000-04-11
; PCT APPLICATION NUMBER: PCT/JP98/04796
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: JP 9/290188
; PRIOR FILING DATE: 1997-10-22
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 1569
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1557)

```

Alignment Scores:	
Pred. No.:	4.87e-249
Score:	2580.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	4
Length:	1569
Matches:	504
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-830-144-4 (1-504) x US-09-529-279-10 (1-1569)

Qy	1	MetAlaAlaGlnArgSerLeuLeuGlnSerGluGlnGlnProSerTrpThrAspAsp	20	Qy	361	LeuValArgAsnPhGlyTyrProLeuGlyGluMetSerGlnProThrProSerProAla	380
Db	7	ATGCGCGCCAGAGGAGGAGCTTGCTGCAGAGTGAGCAGCAGCAAGCTGGCAGATGAC	66	Db	1087	CTAGTAGAGAACTTTGGCTACCCCTGGCGGAAATGAGCAGGCCACACGAGGCCGACCC	1146
Qy	21	LeuProLeuCysHisLeuSerGlyValGlySerAlaSerAsnArgSerTyrSerAlaAsp	40	Qy	381	ProAlaLaGlyGlyArgValTyrProValSerValProTyrSerSerAlaGlnSerThr	400
Db	67	CTGCCTCTCTGCCACCTCTCTGGGGTTGGCTCAGCCCTCCAACCGCAGCTACTCTGCTGAT	126	Db	1147	CCAGCTGCAGAGAGAGCAGGTACCTTGCTGTCTGTGCCTACTCCAGCGCCGACAGCACC	1206
Qy	41	GlyLysGlyThrGluSerHisProProGluAspSerTrpLeuLysPheArgSerGluAsn	60	Qy	401	SerLysThrSerValThrLeuSerLeuValMetProSerGlnGlyGlnMetValAsnGly	420
				Db	1207	AGCAAGACCCAGCGTGACCTCTCCCTTGTCATGTGCCCCAGGCGCCAGATGCTCAACGGG	1266

QY 421 AlaHisSerAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThr 440  
Db 1267 GCTACAGTCTTCACCCCTGGAGAGCAACCCCTTCACCAACCAAGCCGAGCC 1326  
QY 441 LeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerAspGly 460  
Db 1327 TTAACTGTGACGTCCACCAACAGCAGCAGCAGCTCCAGCTCTGAACGAGGC 1386  
QY 461 LeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGlyAlaGluPro 480  
Db 1387 CTCTTCGCTCCCGCCGCCCTCCCTCCGCTGCGCAGAGCAGCTCGTGTGAGCCC 1446  
QY 481 TyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGlyGlnSerVal 500  
Db 1447 TATGTGACCTTGCTGAGTTTACCGCTTGAGCGTGAGCACTGCGAGCAGCGCTG 1506  
QY 501 ValThrAlaPro 504  
Db 1507 GTGACAGCACCG 1518

RESULT 6  
US-09-529-279-42  
; Sequence 42, Application US/09529279  
; Patent No. 6451617  
; GENERAL INFORMATION:  
; APPLICANT: ONO, KOICHIRO  
; APPLICANT: OHTOMO, TOSHITAKO  
; APPLICANT: TSUCHIYA, MASAYUKI  
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES  
; FILE REFERENCE: 053466/0278  
; CURRENT APPLICATION NUMBER: US/09/529, 279  
; CURRENT FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: PCT/JP98/04796  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: JP 9/290188  
; PRIOR FILING DATE: 1997-10-22  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 42  
; LENGTH: 1568  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (11) .. (1549)  
US-09-529-279-42

Alignment Scores:  
Pred. No.: 9.72e-249 Length: 1568  
Score: 2577.00 Matches: 503  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.80% Mismatches: 0  
Query Match: 99.88% Indels: 0  
Gaps: 0

US-09-830-144-4 (1-504) x US-09-529-279-42 (1-1568)

QY 1 MetaAlaGlnArgSerLeuGlnSerGlnGlnGlnProSerTrpThrAsp 20  
Db 38 ATGGCGGCGCAGAGAGAGGAGCTTGCTGACAGATGAGCAGCAGCCAACTGAGCAGATAC 97  
QY 21 LeuProLeuCysHisLeuSerGlyValGlySerAlaSerAsnArgSerTyrSerAla 40  
Db 98 CTGCTCTCTGACCTCTCTGAGGTGAGCTGAGCTCCAAACCGAGCTACTCTGCTAAT 157  
QY 41 GlyIysGlyThrGlnSerHisProGluAspSerTrpLeuLeuPheArgSerGluAsn 60  
Db 158 GCGAAGGAGCAGAGAGCAGCAGCCGCGCAGAGAGCAGCTGCTCAAGTTCAGAGTGAAG 217  
QY 61 AsnCysPheLeuTyrGlyValPheAsnGlyTyrAspGlyAsnArgValThrAsn 80  
Db 218 AACTGCTCTCTGATGAGGCTTTTCAACGCGCTATGATGAGCAACCGAGTGAACCACTTCG 277

QY 81 AlaGlnArgLeuSerAlaGluLeuLeuGlyGlnLeuAsnAlaGlnHisAlaGluAla 100  
Db 278 GCCCAGCGGCTGTGCTGCGAGAGACTCTGCTGGCCAGCTGAATGCCAGCAGCGAGGCC 337  
QY 101 AspValAlaArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGlnSer 120  
Db 338 GATGTGCGCGGTGTGCTGCTGAGGCTTCATGTGTGGAGAGAGTTCCTGAGAGTCC 397  
QY 121 IleAspAspAlaLeuAlaGluLeuAlaSerLeuGlnSerGlnLeuProGluGlyValPro 140  
Db 398 ATTGACAGAGCTTGCTGAGAGAGCAAGCTCCAGTGGCAATGCCAGAGGAGTCCCT 457  
QY 141 GlnHisGlnLeuProGlnTyrGlnLysIleLeuGluArgLeuLeuValArg 160  
Db 458 CAGCAGCAGAGCTGCTCCCTCAGATCAGAAAGATCCTTGAGACATCCAAAGCCTTGAAGAG 517  
QY 161 GluIleSerGlyGlyAlaMetAlaValAlaValLeuLeuAsnAsnLysLeuTyrVal 180  
Db 518 GAAATTTCCGAGGCGCATGCGCTGTGGCGGTCTTCACCAACACCTTACGTC 577  
QY 181 AlaAsnValGlyThrAsnArgAlaLeuLeuCysLysSerThrValAspGlyLeuGlnVal 200  
Db 578 GCCAATGTGTTACAAACCGTGCACTTTATGCAAATGCACAGTGGATGGTTGAGGTG 637  
QY 201 ThrGlnLeuAsnValAspHisThrThrGluAsnGluAspGluLeuPheArgLeuSerGln 220  
Db 638 ACACAGCTGAAGTGGACCAACCAACAGAGAGAGTGAAGTCTTCGCTTCGCGAG 697  
QY 221 LeuGlyLeuAspAlaGlyLysIleLysGlnValGlyIleLysGlyGlnGlnSerThr 240  
Db 698 CTGGCGCTTGATGTGAGAAAGATCAAGCAGGTGGGATCATCTGTGGCAGAGAGCACCC 757  
QY 241 ArgArgIleGlyAspTyrLysValLysTyrGlyTyrThrAspIleAspLeuSerAla 260  
Db 758 CGCGGATCGGGATTCAAAGTTAAATGTGCTACACGAGATTCCTTCAGCGCT 817  
QY 261 AlaLysSerLysProIleLeuAlaGluProGluIleHisGlyAlaGlnProLeuAspGly 280  
Db 818 GCCAATCCAAACCAATCATCCAGAGCAGAAATCCATGGGGCAGCGCTGATGG 877  
QY 281 ValThrGlyPheLeuValLeuMetSerGlyGlyLeuTyrLysAlaLeuGlnAlaHis 300  
Db 878 GTGACGGCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 937  
QY 301 GlyProGlyGlnAlaAsnGlnGluIleAlaMetIleAspThrGluPheAlaLysGln 320  
Db 938 GGGCTGGGAGGCCCAACAGAGATTCCTGCGATGATTCGACATGAGTTGCCAAGCAG 997  
QY 321 ThrSerLeuAspAlaValAlaGlnAlaValAlaAspArgValLysArgIleHisSerAsp 340  
Db 998 ACCTCCCTGGACGCAATGGCCGCGCGTGTGAGACCGGGTGAAGGAGTCCACAGCGCAG 1057  
QY 341 ThrPheAlaSerGlyGlyLysArgAlaArgPheCysProArgHisGluAspMetThrLeu 360  
Db 1058 ACCCTGCGCAGGTGTGGGAGCGTGCAGAGTTCTGCCCCGCGACAGAGACATGACCTTG 1117  
QY 361 LeuValArgAsnPheGlyTyrProLeuGlyGluMetSerGlnProThrProSerProAla 380  
Db 1118 CTAGTGAAGAACTTTGGCTTACCCGCTGGGCCCAATAGCAGCGCCACAGCCAGCCAGCC 1177  
QY 381 ProAlaAlaGlyValArgValTyrProValSerValProTyrSerSerAlaGlnSerThr 400  
Db 1178 CCAAGTCAAGAGAGAGAGAGTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1237  
QY 401 SerLysThrSerValThrLeuSerLeuValMetProSerGlnGlyGlnMetValAsnGly 420  
Db 1238 AGCAAGACCAAGCCTGACCTTCTCTTGTATGCTCTCCAGGGGCGAGATGGTCAAGGAG 1297  
QY 421 AlaHisSerAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThr 440  
Db 1298 GCTACAGTCTTCCACCTTCAGAGAGAGCAGCAGCCCACTTACCAACCAAGCCCGAGCC 1357



QY 341 ThrPheAlaSerGIyGIyAluAgaAlaArgPheCysProArgHisGluAspMetThrLeu 360  
Db 1050 ACCTTCCGCGAGTGGGGAGCGCGCCAGGTTCTGCCCGGACCGAGGACATGACCTTG 1109  
QY 361 LeuValArgAsnPhgGIyTYrProLeuGIyGluMetSerGlnProThrProSerProAla 380  
Db 1110 CTAGTGAAGAACTTGGCTTACCCCTGGGCGAATGAGCCAGCCCAACCGAGCCGACGCC 1169  
QY 381 ProAlaAlaGIyGIyArgValTYrProValSerValProTYrSerSerAlaGlnSerThr 400  
Db 1170 CCAGCTGAGAGAGGAGGAGGAGTACCTGTCTGTGCATCTCAGCGCCCAAGACACC 1229  
QY 401 SerTYrThrSerValThrLeuSerLeuValMetProSerGlnGluMetValAsnGIy 420  
Db 1230 AGCAAGACCGACGTGACCTCTCCCTGTGCATGCCCTCCGAGGCGCAGATGTAAAGGG 1289  
QY 421 AlaHisSerAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThr 440  
Db 1290 GCTCACAGTGTCTTCCACCTTGAGCAGAGCACCACCTCACCACCAAGCCGACACC 1349  
QY 441 LeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerSerAspGIyGIy 460  
Db 1350 TTAACCTCGACGTCCACCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1409  
QY 461 LeuPheArgSerArgProAlaHisSerLeuProProGIyGIyAspGIyArgValGluPro 480  
Db 1410 CTCTTCCGCTCCCGGCGCCGCTGCTCCGCTGCGAGAGAGGTGGTGTGAGGCC 1469  
QY 481 TYrValAspPheAlaGluPheTYrArgLeuTYrSerValAspHisGluGlnSerVal 500  
Db 1470 TATGTGAGACTTGTGAGTTTATACCGCTCTGAGCGTGAGACCAATGGAGAGAGCCTG 1529  
QY 501 ValThrAlaPro 504  
Db 1530 GTGACAGCACCG 1541

RESULT 8  
US-09-144-178-5  
Sequence 5, Application US/09144178  
Patent No. 5989862  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihiko  
APPLICANT: NISHIDA, Eisuke  
TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/144,178  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/752,891  
FILING DATE: 20-NOV-1996  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1560 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1541  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 30..1541  
US-09-144-178-5

Alignment Scores:  
Pred. No.: 1,53e-248 Length: 1560  
Score: 2575.00 Matches: 503  
Percent Similarity: 99.80% Conservative: 0  
Best Local Similarity: 99.80% Mismatches: 1  
Query Match: 99.81% Indels: 0  
Gaps: 0

US-09-830-144-4 (1-504) X US-09-144-178-5 (1-1560)

QY 1 MetAlaAlaGlnArgArgSerLeuLeuGlnSerGlnGlnProSerThrThrAsp 20  
Db 30 ATGGCGCGGAGAGGAGGAGGAGCTTGTGCAGAGTGAACAGCAGCAGCAGCAGCAGTGC 89  
QY 21 LeuProLeuCysHisLeuSerGIyValGIySerAlaSerAsnArgSerTYrSerAlaAsp 40  
Db 90 CTGCTCTTCTGCCACCTCTCTGGGGTTGGCTCAGCTCCACCGCAGCTACTGCTGAT 149  
QY 41 GlyLeuGIyThrGlnSerHisProProGluAspSerTYrLeuLeuPheArgSerGluAsn 60  
Db 150 GGCAAGGGCACTGAGAGCCACCAGCAGAGCAGATGGCTCAAGTTCAGAGTGAAGAC 209  
QY 61 AsnCysPheLeuTYrGIyValPheAsnGIyTYrAspGIyAsnArgValThrAsnPheVal 80  
Db 210 AACTGCTTCTGATGGAGGCTTTCACGSGCTATGATGCAACCGATGACCACTTCGTG 269  
QY 81 AlaGlnArgLeuSerAlaGluLeuLeuGluGlnLeuAsnAlaGluHisAlaGluAla 100  
Db 270 GCCAAGCGGCTGTCCGAGAGCTTCCTGGCCAGCTGAATCCGAGCAGCCGAGAGCC 329  
QY 101 AspValArgArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGlnSer 120  
Db 330 GATGTGGCGCGTGTGCTGCTGACAGCCCTTGATGTGTGAGAGAGACTTCTGGAGTCC 389  
QY 121 IleAspAsnAlaLeuAlaGluValSerLeuGlnSerGlnLeuProGluGIyValPro 140  
Db 390 ATTACGAGCGCTTGGCTGAGAGGCAAGCTCCAGTCCGCAATTTGCCAGAGGAGTCCCT 449  
QY 141 GlnHisGlnLeuProProGlnTYrGlnHisIleLeuGluArgLeuTYrThrLeuGluArg 160  
Db 450 CAGCACCGAGCTGCTCCCTCACTATCAGAAATCTTGAAGACTCAAGAGCTTGAAGAGG 509  
QY 161 GluIleSerGIyGIyAlaMetAlaValValAlaValLeuLeuAsnLeuValSerTYrVal 180  
Db 510 GAAATTTGGGAGGGGCGCATGGCGGTGTGCTTCAACAACAACAGCTTACGTC 569  
QY 181 AlaAsnValGIyTYrAsnArgAlaLeuLeuCysIysSerThrValAspGIyLeuGlnVal 200  
Db 570 GCCAATGCTCGTACCAACCGTGCATTTTATGCAATGACAGTGGATGGGTTCAAGTGC 629  
QY 201 ThrGlnLeuAsnValAspHisThrThrGluAsnGluAspGIyLeuPheArgLeuSerGln 220  
Db 630 ACAGAGCTGAAGTGAACCAACACACAGAGAGATGAGTCTTCCGCTTTTCGACG 689



```
QY 221 LeuGlyLeuAspAlaGlyLysIleLysGlnValGlyIleIleCysGlyGlnGluSerThr 240
DB 690 CTGGGCTTGGATCTTGGAAAGATCAAGCAGGTGGGATCATCTGTGGCGAGAGCACC 749
QY 241 ArgArgIleGlyAspTyrLysValLysTyrGlyTyrThrAspIleAspLeuLeuSerAla 260
DB 750 CGCGGATCGGGGATTACAAGGTTAAATATGCTACAGGACATTGACCTTCTCAGCGCT 809
QY 261 AlaLysSerLysProIleIleAlaGluProGluIleHisGlyAlaGlnProLeuAspGly 280
DB 810 GCCAAGTCCAAACCAATCATCGAGAGCCAGAAATCCATGGGCACAGCCCTGGATGG 869
QY 281 ValThrGlyPheLeuValLeuMetSerGluGlyLeuTyrLysAlaLeuGluAlaHis 300
DB 870 GTGACGGGCTTCTGGTGTGATGTCGAGGGGTGTACAGGCCCTAGAGGAGCCCAT 929
QY 301 GlyProGlyGlnAlaAsnGlnGluIleAlaAlaMetIleAspThrGluPheAlaLysGln 320
DB 930 GGGCTGGCAGGCCAACCCAGGAGATTGCTCGGATGATTGACACTGAGTTTGCACGAC 989
QY 321 ThrSerLeuAspAlaValAlaGlnAlaValAlaAspArgValLysArgIleHisSerAsp 340
DB 990 ACCTCCCTGGACGAGTGGCCAGGCCGCTCGTGACCGGTGAAGCGCATCCACAGCGAC 1049
QY 341 ThrPheAlaSerGlyGlyGluArgAlaArgPheCysProArgHisGluAspMetThrLeu 360
DB 1050 ACCTTCGCCAGTGTGGGAGCGTGCAGGTTCTGCCCCCGGCACGAGACATGACCCTG 1109
QY 361 LeuValArgAsnPheGlyTyrProLeuGlyGluMetSerGlnProThrProSerProAla 380
DB 1110 CTAGTGAGGAACCTTTGGCTACCCGCTGGCGGAAATGAGCCAGCCACCCGAGCCAGCC 1169
QY 381 ProAlaAlaGlyArgValTyrProValSerValProValSerTyrSerSerAlaGlnSerThr 400
DB 1170 CCAGCTCGAGAGGACAGTGTACCTGTGTGTGTCATCTCCAGCTCCAGGCCCGAGAGCACC 1229
QY 401 SerLysThrSerValThrLeuSerLeuValMetProSerGlnGlyGlnMetValAsnGly 420
DB 1230 AGCAAGACCGAGTGCCTCTCCCTGTGTCATGCTCCAGGCCAGATGTCACCGGG 1289
QY 421 AlaHisSerAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThr 440
DB 1290 GCTCACAGTCTCCACCTTGGACGAGCCACCCACCCCTCACCAACCAAGCCGAC 1349
QY 441 LeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerSerAspGlyGly 460
DB 1350 TTAACCTCGAGTCCACCAACAGCCACAGCAGCAGCAGCAGCTCCAGCTCTCAGCGAGGC 1409
QY 461 LeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGlyArgValGluPro 480
DB 1410 CTCTTCGCTCCCGGCCCGCCACTCGCTCCCGCTGGCGAGGACGCTCGTTGAGGCC 1469
QY 481 TyrValAspPheAlaGluPheTyrArgLeuTyrSerValAspHisGlyGlnSerVal 500
DB 1470 TATGTGACTTTTCTGAGTTTTTACCAGCTTCTGGAGCGTGGACCATGGCGAGCAGAGCGTG 1529
QY 501 ValThrAlaPro 504
DB 1530 GTGACAGCACCG 1541
```

## RESULT 9

US-09-406-854-5

; Sequence 5, Application US/09406854

; Patent No. 6140042

; GENERAL INFORMATION:

; APPLICANT: MATSUMOTO, Kunihiro

; APPLICANT: NISHIDA, Eisuke

; TITLE OF INVENTION: TABI PROTEIN AND DNA CODING THEREFOR

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley &amp; Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

```
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/406,854
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/752,891
FILING DATE: 20-NOV-1996
APPLICATION NUMBER: JP 8-300856
FILING DATE: 28-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-126282
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17981/111
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1560 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1541
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 30..1541
US-09-406-854-5
Alignment Scores:
Pred. No.: 1.53e-248
Score: 2575.00
Percent Similarity: 99.80%
Best Local Similarity: 99.80%
Query Match: 99.81%
DB: 3
Length: 1560
Matches: 503
Conservative: 0
Mismatch: 1
Indels: 0
Gaps: 0
US-09-830-144-4 (1-504) x US-09-406-854-5 (1-1560)
QY 1 MetaAlaAlaGlnArgSerLeuLeuGlnSerGluGlnGlnProSerTrpThrAspAsp 20
DB 30 ATGGCGGCGCAGAGGAGGAGCTTCTCAGAGTGAGCAGCAGCAGCAGCTGGACAGATGAC 89
QY 21 LeuProLeuCysHisLeuSerGlyValGlySerAlaSerAsnArgSerTyrSerAlaAsp 40
DB 90 CTGCCCTCTGCCACCTCTCTGGGGTTGGCTCAGCCTCCAAACCGCAGCTACTCTGCTGAT 149
QY 41 GlyLysGlyThrGluSerHisProGluAspSerTrpLeuLysPheArgSerGluAsn 60
DB 150 GGCAGGGCAGCTGAGAGCCACCCCGCAGAGCAGATGGGCTCAAGTTCAGGAGTGAGAAC 209
QY 61 AenCysPheLeuTyrGlyValPheAsnGlyTyrAspGlyAsnArgValThrAsnPheVal 80
DB 210 AACTGCTTCTGTATGGGTTCTTCAACGGCTATGATGGCAACCGAGTGACCACTTCGTG 269
QY 81 AlaGlnArgLeuSerAlaGluLeuLeuGlyGlnLeuAsnAlaGluHisAlaGluAla 100
DB 270 GCCCAGCGGCTGTCCGAGAGCTCTCTGCGGCAGCTGAATGCCGAGCAGCCGAGGCC 329
QY 101 AspValArgArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGluSer 120
```



Db	330	GATGTGGCGCGTGTCTGCTGCTGAGGCTTTGATGTGGTAGAGGAGCTTCTCGAGTCC	389
OY	121	lLeaPaSPaAlaLeuAlaGluLeAlaSerLeuGlnSerGlnLeuProGluGlyValPro	140
Db	390	ATTGACGACGCGCTTGGCTGTGAGAAAGGACCTCCAGTCCGAAATTGCCAGAGGAGTCCCT	449
OY	141	GlnHsGlnLeuProProGlnlntTyGlnLysellLeuGlnLArgLeuLysThyLeuGlnArg	160
Db	450	CAGACCCAGCGCTGCTCTCAATATCAAGAAATCCTTGAGAGATCTCAAGACGTTAGAGAGG	509
OY	161	GluLLeSerGlyGlyAlaMetAlaValValAlaValLeuLysAsnLysLeuTyAla	180
Db	510	GAATTTTGGAGAGGGCCATGCGCGTTGTGCGGCTCTTCAACAACAGCTCTACGTC	569
OY	181	AlaAsnValGlyThyAsnArgAlaLeuLeuCysLysSerThyValAspGlyLeuGlnVal	200
Db	570	GCCATGTGCGTATCAAAACCGTGACCTTTATGCAATTCGACATGGATGGGTTCGACGTG	629
OY	201	ThyGlnLeuAsnValAspHisLThrThyGlnAsnGlnAspGlnLeuPheArgLeuSerGln	220
Db	630	ACACAGCTGAACCGTGAGCCACACACAGAAACAGAGATGAGCTCTTCGCTTCGACG	689
OY	221	LeuGlyLeuAspAlaGlyLysLLeLysGlnValGlyLLeLLeGlyGlyGlnGlnSerThy	240
Db	690	CTGGGCTTGGATGCTGGAAGAAATCAACAGAGTGGGATCATCTGTGGGAGAGAGACAC	749
OY	241	ArgArgLLeGlyAspTyTyLysValLysTyGlyTyTyThyAspLLeAspLeuSerAla	260
Db	750	CGCGCGAGTCGGGAGTTACAGAGTTAAATATAGCTACACGACATTACCTTCTCAGCGCT	809
OY	261	AlaLysSerLysProLLeLLeAlaGlnProGlnLLeHisGlyAlaGlnProLeuAspGly	280
Db	810	GCCAAAGTCAAAACCAATCATGCGACAGCCAGAAATCCATGGGACAGCGCTGGATGGG	869
OY	281	ValThyGlyPheLeuValLeuMetSerGlnGlyLeuTyTyLysAlaLeuGlnAlaHis	300
Db	870	GTGAGCGGCTTCTTGGTGCTGATGTCGAGGGGTGTACAAAGCCCTTAAGGACAGCCAT	929
OY	301	GlyProGlyGlnAlaAsnGlnGlnLLeAlaAlaMetLLeAspThyGlnPheAlaLysGln	320
Db	930	GGGCTGGGCGGCGCAACACAGAGATTGCTGCATGATGTACCTGAGTTTCCAAAGAC	989
OY	321	ThySerLeuAspAlaValAlaGlnAlaValValAspArgValLysArgLLeHisSerAsp	340
Db	990	ACCTCCCTGAGCGGAGTGGCGGCGCGCTGTGACCGGGTGAAGGCATCCACAGCAC	1049
OY	341	ThyPheAlaSerGlyGlyGlnArgAlaArgPheCysProArgHisGlnAspMetThyLeu	360
Db	1050	ACCTTCGCAATGTGTGGGAGCGTGACGAGTTTCTGCCCGGACGAGACATGACCTCG	1109
OY	361	LeuValArgAsnPheGlyTyTyProLeuGlyGlnMetSerGlnProThyProSerProAla	380
Db	1110	CTAGTAGAGAACTTTGGCTACCGGCTGTGGGAATAGCCAGCCACACCGAGGCCACCC	1169
OY	381	ProAlaAlaGlyGlyArgValTyTyProValSerValProTySerSerLLeGlnSerThy	400
Db	1170	CCAGCTGAGGAGAGACGAGTGTACCTCTGTCTGTGCATATCCACGGCCCAAGACACC	1229
OY	401	SerLysThySerValThyLeuSerLeuValMetProSerGlnGlyGlnMetValAsnGly	420
Db	1230	AGCAAGACACAGCCGACCTCTCCCTTGTATGCTCCCTCCAGGCGCAAGTGTCAACGG	1289
OY	421	AlaHisSerAlaSerThyThyLeuAspGlnAlaLThrProThyLeuThyAsnGlnSerProThy	440
Db	1290	GCTCACTAGTCTTCACCTCGGAGAGAGCCACCCACCTCAACCAACCAAGCCCGAC	1349
OY	441	LeuThyLeuGlnSerThyAsnThyHisThyGlnSerSerSerSerSerAspGlyGly	460
Db	1350	TTAACCTCGCAGTCCACCAACAGCCACGCAAGCAGCAGCTCCAGCTCTACGAGGAGCC	1409
OY	461	LeuPheArgSerArgProAlaHisSerLeuProProGlyGlnAspGlyArgValGlnPro	480

```

Db      1410  CTCTTCGCCGCTCCGGGCCCGCCCACTGCTCCGCTGGCGAGAGAGCGTGCTTTAGCC 1463
QY      481  TyrValAspPhealaIagIuPheTyrATgLeuTTPSerValAspHisGIYQIuGInSerVal 500
Db      1470  TATGTGAGACTTGTGTGATGTTTACCGGCTCTGTGAGCGTGACCAATGGGAGAGAGCGCTG 1529
QY      501  ValThrAlaPro 504
Db      1530  GTGACAGCACCG 1541
RESULT 10
US-09-013-881-10
/ Sequence 10, Application US/09013881
/ Patent No. 6132964
/ GENERAL INFORMATION:
/ APPLICANT: Bandman, Olga
/ APPLICANT: Lal, Preeti
/ APPLICANT: Hillman, Jennifer L.
/ APPLICANT: Corley, Neil C.
/ APPLICANT: Guegler, Karl J.
/ APPLICANT: Shah, Puryi
/ TITLE OF INVENTION: HUMAN HYDROLASE-LIKE MOLECULES
/ NUMBER OF SEQUENCES: 16
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Incyte Pharmaceuticals, Inc.
/ STREET: 3174 Porter Drive
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/013,881
/ FILING DATE: HEREWITH
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BILLINGS, LUCY J.
/ REGISTRATION NUMBER: 36,749
/ REFERENCE/DOCKET NUMBER: PF-0470 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650-855-0555
/ TELEFAX: 650-845-4166
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1403 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: KIDNOT02
/ CLONE: 195647
/ US-09-013-881-10
Alignment Scores:
Pred. No.: 2,77e-08 Length: 1403
Score: 178.00 Matches: 83
Percent Similarity: 40.88 Conservative: 65
Best Local Similarity: 22.93 Mismatches: 116
Query Match: 6.90% Indels: 98
DB: Gaps: 18
US-09-830-144-4 (1-504) x US-09-013-881-10 (1-1403)
QY      3  AlaGlnArgSerLeuLeuGlnSerGluGlnGlnProSerTTPThrAspAspLeu--- 21

```

```
Db 302 GCAAGAGAAAACC-----TCCGAGGAAGAGAAATGCGCAGTGAAGAGCTTGTG 352
Qy 22 -----ProLeuCysHisLeuSerGlyValGlySerAlaSerAsnArgSerTyrSerAla 39
Db 353 GAAAGAAAGTTTGAAGACCTCTTCGGTG---ATCTTTGGTCTGAAGGGCTATGTGCT 409
Qy 40 AspGlyValGly-----Thr 44
Db 410 GAGCGGAAGGGTGAGAGGAGAGATGCGAGGATGCCAGCTCATCTGAACGACATCAC 469
Qy 45 GluSerHisProProGluAspSerTrpLeuLysPheArgSerGluAsnAsnCysPheLeu 64
Db 470 GAGGAGTAGGCCCCATCGTCCCTCAT-----ACTCGGGTTTCATAT 514
Qy 65 TyrGlyValPheAsnGlyTyrAspGlyAsnArgValThrAsnPheValAlaGlnArgLeu 84
Db 515 TTTGCTGTTTTTGATGCATCGAGAAATTCGAGCCTCAAAATTTGCTGCACAGAAATTG 574
Qy 85 SerAlaGluLeu-----LeuGlyGlnLeuAsnAlaGluHisAlaGluAlaAspVal 102
Db 575 CATCAAACTTAATCAGAAAATTTCTTAAGGAGATGTAATCAGTGTAGAGAAACCGTG 634
Qy 103 ArgArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGluSerIleAsp 122
Db 635 AAGAGATGCTTTTGGACACTTTCAGCATACTGATGAGAGATTCTT----- 682
Qy 123 AspAlaLeuAlaGluLysAlaSerLeuGlnSerGlnLeuProGluGlyValProGlnHis 142
Db 683 -----AAACAAGCTTCAGCCAGAG----- 703
Qy 143 GlnLeuProProGlnTyrGlnLysIleLeuGluArgLeuLysThrLeuGluArgGluIle 162
Db 704 -----CCTGCCCTGGAAA----- 715
Qy 163 SerGlyGlyAlaMetAlaValAlaValAlaLeuLeuAsnAsnLysLeuTyrValAlaAsn 182
Db 716 ---GATGGTCCACTCCACGCTGTCTTGGCTGTAGACAACTTTATATTGCCAAC 772
Qy 183 ValGlyThrAsnArgAlaLeuLeuCysLysSerThrValAspGlyLeuGln-----Val 200
Db 773 CTCGGAGATAGTCGGGCAATCTGTGCTGTATATGAGGAGAGTCAAAACATCGAGCC 832
Qy 201 ThrGlnLeuAsnValAspHisThrThrGluAsnGluAspGluLeuPheArgLeuSerGln 220
Db 833 TTAAGCCTCAGCAAGAGCATATCAACTCATGATGAAGACGGATGAGGATACAGAG 892
Qy 221 LeuGlyLeuAspAlaGlyLysIleLysGlnValGlyIleIleCysGlyGlnGluSerThr 240
Db 893 CTGGA-----CGAAACGTGAGGATGGCGTGTTCGGCGTGTAGAGGTGTCA 943
Qy 241 ArgArgIleGlyAsp-----TyrLysValLysTyrGlyTyrThrAspIleAspLeuLeu 258
Db 944 CGCTCCATTGGGACGGCAGTACAAG---CGTGGCGGTGCACCTCTGTG----- 991
Qy 259 SerAlaAlaLysSerLysProIleIleAlaGluProGluIleHisGlyAlaGln----- 276
Db 992 -----CCGACATCAGACGCTGCCAGCTGAC 1018
Qy 277 ProLeuAspGlyValThrGlyPheLeuValLeuMetSerGluGlyLeuTyrLysAlaLeu 296
Db 1019 CCAATGAC-----AGGTTCAATTGTGGCCCTGTGATGGGCTCTTCAAGGCTTT 1069
Qy 297 GluAlaAlaHisGlyProGlyGlnAlaAsnGlnGluIleAlaAlaMetIleAspThrGlu 316
Db 1070 ACC-----CCAGAAGAAGCCGTGAACCTTCTTGTCTGTCTCGAGGATGAA 1117
Qy 317 PheAlaLys-----GlnThrSerLeuAspAlaValAlaGlnAlaValValAsp 332
Db 1118 AAGATCCAGACCCGGGAAGGAAGTCGACAGCCGCGCTACGAAGCAGCCTGCAAC 1177
Qy 333 ArgVal 334
Db 1178 AGGCTG 1183
```

```
RESULT 11
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezana
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Sorangium cellulosum
; IMMEDIATE SOURCE:
; CLONE: p98/1, pJL3, and pVKM15
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 383..760
; OTHER INFORMATION: /product= "SorR"
; OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous
; OTHER INFORMATION: the reductase domains of type I PKs such as eryA from
; OTHER INFORMATION: Saccharopolyspora erythraea."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 927..19874
; OTHER INFORMATION: /product= "SorA"
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs that
; OTHER INFORMATION: are known to be involved in the synthesis of polyketide
; OTHER INFORMATION: compounds."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 942..7115
; OTHER INFORMATION: /product= "Module 1 of SorA"
; FEATURE:
```

```

/ NAME/KEY: misc feature
/ LOCATION: 7203..12884
/ OTHER INFORMATION: /product= "Module 2 of SorA"
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 13455..19616
/ OTHER INFORMATION: /product= "Module 3 of SorA"
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 19871..46318
/ OTHER INFORMATION: /product= "SorB"
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 19870..24556
/ OTHER INFORMATION: /product= "Module 1 of SorB"
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 24638..30820
/ OTHER INFORMATION: /product= "Module 2 of SorB"
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 30881..35446
/ OTHER INFORMATION: /product= "Module 3 of SorB"
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 35528..40114
/ OTHER INFORMATION: /product= "Module 4 of SorB"
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 40190..46318
/ OTHER INFORMATION: /product= "Module 5 of SorB"
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 46851..47891
/ OTHER INFORMATION: /product= "SorM"
/
/ OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly
/ OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
/ OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
/ OTHER INFORMATION: polyketide rapamycin."
/
US-08-764-233A-1

Alignment Scores:
Pred. No.: 0.487
Score: 131.00
Percent Similarity: 33.56%
Best Local Similarity: 22.76%
Query Match: 5.08%
DB: 1
Gaps: 19

US-09-830-144-4 (1-504) x US-08-764-233A-1 (1-49377)
QY 84 LeuSerAlaGluLeuLeuLeuGlyGlnLeuAsnAlaGluHisAlaGluAlaAspValArg 103
Db 9057 CTGAAGCCCGCTCTCTCTCGGACACATCCATGCGGAGCTCTGCGGCCCGACGTGCC 9116
QY 104 ArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGluSerIleAspAsp 123
Db 9117 GCGGCTCTT-----TCTCTCCAGGAC 9137
QY 124 Ala-----LeuAlaGluLysAlaSerLeuGlnSerGlnLeuProGluGlyValPro 140
Db 9138 GCGTCACCCCTCTGTCGCGCCGCAAGCTCATGCAAGGCGCTCCCAAGAGGCGGCC 9197
QY 141 GlnHisGlnLeuProProGlnTyrGlnLysIleLeuGluArgLysThrLeuGluArg 160
Db 9198 ATGTCAACCTTCGAGCTCCGAGAGAGAAATCGCGACCTTCTCCAGCCCTACGATGA 9257
QY 161 GluIleSerGlyGlyAlaMetAlaValValAlaValLeuLeuAsnLysLeuTyrVal 180
Db 9258 CGAGTACCTTCGCGCC-----CTCAATGGGCTCTCTCCACC 9296
QY 181 AlaAsnValGlyThrAsnArgAlaLeuLeu-----CysLysSerThrValAspGly 197

```

```

Db 9297 GTCTGCTGCGCATGAAGACGCGGTGTGGATGCGCCCGACGCGCAAGCCCTCGGA 9356
QY 198 LeuGlnValThrGlnLeuAsnValAspHisThrThrGluAsnGlu----- 212
Db 9357 CGAAGACCAACACGCTGCGCGGTGAGCGACCTTCACCTCCGACATGACGGAATG 9416
QY 213 --AspGluLeuPheAspArgLeuSerGln--LeuGlyLeuAspAlaGlyIleLysGln 230
Db 9417 CTCGACGACCTTCGCGCGGTGCGCCAGAGCTTACCTACCATCCCGACGATCCCATC 9476
QY 231 ValGlyIleIleCysGlyGlnLysSerThrArgArgIleGlyAspTyrLysValLysTyr 250
Db 9477 ATCTCAACGTCACCGCGCGCGCGCACG-----GACCACGAG----- 9515
QY 251 GlyTyrThrAspIleAspLeuLeuSerAlaAlaLysSerLysProIleIleAlaGluPro 270
Db 9516 -----CTGCGCTGCGCC 9527
QY 271 GluIle-----HisGlyAlaGlnProLeuAspGlyValThrGlyPhe 284
Db 9528 GACTACTGGGTCCGCCACGCTTCCGACACCGCTCCCTTCTCCAGCGCGTA----- 9578
QY 285 LeuValLeuMetSerGluGlyLeuTyrLysAlaLeuGluAlaAlaHisGlyProGlyGln 304
Db 9579 -----CGTGCCTTCACGCGCAAGG--- 9599
QY 305 AlaAsnGlnGluIleAlaAlaMetIleAspThrGluPheAlaLysGlnThrSerLeuAsp 324
Db 9600 -----GCACGCGCTTTCTCGACCTCGGCGCTCACGCTGCTCTCC 9641
QY 325 AlaValAlaGlnAlaValVal--AspArgValLys--ArgIleHisSerAspThrPheAl 343
Db 9642 GCGCTTCGCAAGACGCGCTTCGACACAGACGAAGGACGCTCGCATTCGCTTCTCC 9701
QY 343 AspGlyGlyGluArgAla-----ArgPheCysProArgHisGluAspMetThr 359
Db 9702 ACCCTCCGCAAGGACGCGCGACGCGGAGCGTTCACGCGCGCGTGGCGCTTCAC 9761
QY 359 fLeuLeuValArgAsnPheGlyTyrProLeuGlyGluMetSerGln----- 374
Db 9762 GCTG-----CAGGCTCACACCTGACTGAGGCGCTTCTTCGCCCCCTTCCTCCA 9812
QY 375 -----ProThrProSerProAlaProAlaAlaGlyIy----- 385
Db 9813 CGCAAGGTCTCCCTCCCACTATGCTTCCAGCGGAGCGCTTCTGCTCGATGCTTCC 9872
QY 386 ----ArgValTyrProValSerValProTyrSerSerAlaGlnSerThrSerLysThrSe 404
Db 9873 AAGGACACGCTGCGACGCTCGGCTCCGACGCGCTGACCTCGACCGATACCGCGCTGCTC 9932
QY 404 fValThrLeuSerLeuValMetProSerGlnGlnMetValAsnGlyAlaHisSerAl 424
Db 9933 GGGCGCGCGCTCC----- 9945
QY 424 AspThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThrLeuThrLeuGl 444
Db 9946 -----CCCTGCGGACCGCATGGCTTC 9968
QY 444 nSerThr-----AsnThrHisThrGlnSer-----SerSerSerSe 457
Db 9969 CTCTTACAGAGACGACTCTCACTCTCAGAGATCCGTGCTCGCGGATACGCTCGCTTC 10028
QY 457 fAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuPro 471
Db 10029 GGTACACCATCTCTCGGGCACTGCTTCTTCGAGCTTGGCC 10071

RESULT 12
US-08-822-701-1
; Sequence 1, Application us/08822701
; Patent No. 5976853
; GENERAL INFORMATION:
; APPLICANT: Guthridge, Mark

```

APPLICANT: Basilico, Claudio  
TITLE OF INVENTION: NOVEL GROWTH FACTOR INDUCIBLE  
TITLE OF INVENTION: SERINE/THREONINE PHOSPHATASE, FIN13  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
FLOOR  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/822,701  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 1049-1-002 N  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1824 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-822-701-1

Alignment Scores:  
Pred. No.: 0.0039 Length: 1824  
Score: 128.50 Matches: 75  
Percent Similarity: 36.60% Conservative: 52  
Best Local Similarity: 21.61% Mismatches: 129  
Query Match: 4.98% Indels: 91  
DB: 2 Gaps: 12

US-09-830-144-4 (1-504) x US-08-822-701-1 (1-1824)

Qy 165 GlyAlaMetAlaValAlaValLeuLeuAsnLysLeuTyrValAlaAsnValGly 184  
Db 736 GGCACACAGCGGTGGTCTGATCAGAGGGAAGCAGTTGATGGCCAAATGCAGGA 795  
Qy 185 ThrAsnArgAlaLeuLeuCysLysSerThrValAspGlyLeuGlnValThrGlnLeuAsn 204  
Db 796 GACTCTCGCTGTGGTTCGAGGCTGGC-----AAAGCTTAGATATGTC 843  
Qy 205 ValAspHisThrThrGluAsnGluAspGluPheArgLeuSerGlnLeuLeuAsp 224  
Db 844 TATCACCAACACAGAGGATGAAGTGGAGCTGCACGCATCAAGATGCTGGT----- 897  
Qy 225 AlaGlyLysLeuLysGlnValGly---IleIleCysGlyGlnGluSerThrArgGly 243  
Db 898 ---GGCAAGGTACCATGGATGGATGAGTCAATGGAGGCTCAACCTCCAGGGCCATT 954  
Qy 244 GlyAspTyrLysValLysTyrGlyTyrThrAspIleAspLeuLeuSerAlaAlaLysSer 263  
Db 955 GGAGACCACTTCTACAAG-----AGAAAC 978  
Qy 264 LysProIleIleAlaGluProGluIleHisGlyAlaGlnPro----- 277  
Db 979 AAAAATTCGCCACCCAGGAACAGATGATTTCTGCCCTTCTGCATCAAGGTGCTGACT 1038

Qy 278 LeuAspGlyValThrGlyPheLeuValLeuMetSerGluGlyLeuTyrLysAlaLeuGlu 297  
Db 1039 CTACTGATCACCATTGATTCATGCTGCTGTTGACGGGATCTGGAATGTGATG--- 1095  
Qy 298 AlaAlaHisGlyProGlyGlnAlaAsnGlnGluIleAlaAlaMetIleAspThrGluPhe 317  
Db 1096 -----AGCAGCCAGGAGTGTAGACTTTTATTCAATCAAGATC 1134  
Qy 318 AlaLysGlnThrSer-----LeuAspAlaValAlaGlnAlaValVal 331  
Db 1135 AGTCAACGTGATGAAACCGGGAGCTTCGGTTATTGTCTCCATTGTGGAAGAGCTGCTG 1194  
Qy 332 AspArg-----VallysArgIle-Hisse 339  
Db 1195 GATCAGTGCCTGGCCGCAGACACTTCTGGGGATGGTACAGGGTGTGACACATGACGTGC 1254  
Qy 339 rAspThrPheAlaSerGly-----G1 346  
Db 1255 ATCATCATTTGCTTCAAGCCCGAAACACAGTAGAGCTTCAGGCAGAGAGTGGCAAGAG 1314  
Qy 346 yGluArgAlaArgPheCysProArgHisGluAspMetThrLeuLeuValArgAsnPheG1 366  
Db 1315 AAACCTGGAGGAGGACACTGTCCACGGAGGGGGCTGAAGACACCCGGCAACACAGTGA 1374  
Qy 366 yTyrProLeuGlyGluMetSerGlnProThrPro-----SerPr 379  
Db 1375 AAGCCAGAGGAGTACTAGTGGTCAACGGACCTGCCCATGTGACTGTTTCTGAGCCC 1434  
Qy 379 o-AlaPro-----AlaA 383  
Db 1435 TTGACCCCGAGACTGAGTTTGTCTTGTCTTAGCCTTAGCAGTGGGTATGAGGTGTG 1494  
Qy 383 laGlyGlyArgValTyrProValSerValProTyrSerSerAlaGlnSerThrSerLys 403  
Db 1495 CAGGGGCTGGGTGGCTTCTCAGC---CATTAACAAAGGGCCCCCCCCCCCCCA 1551  
Qy 403 hrSerValThrLeuSerLeuValMetProSerGlnGlyGlnMetValAsnGlyAlaHis 423  
Db 1552 CGCGGACCTGGGAGGCTCTGCTGCTCTTAAGCCTCTTACTCTCTCTGAGCTCATC 1611  
Qy 423 erAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThrLeuThrL 443  
Db 1612 GACTATCGTCTCTGCTGCTGCTCTCTGTTGTTGGAGGAAGAGCTGGTAGTTCTGATT 1671  
Qy 443 euGlnSerThrAsnThr 448  
Db 1672 TTTACTCTGTGAACACT 1688

RESULT 13  
US-08-935-855-1  
; Sequence 1, Application US/08935855  
; Patent No. 6066485  
; GENERAL INFORMATION:  
; APPLICANT: Guthridge, Mark  
; APPLICANT: Basilico, Claudio  
; TITLE OF INVENTION: NOVEL GROWTH FACTOR INDUCIBLE  
; TITLE OF INVENTION: SERINE/THREONINE PHOSPHATASE, FIN13  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Jackson, Esq.  
; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
; STREET: Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/935,855

FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 1049-1-002 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1824 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
US-08-935-855-1

Alignment Scores:  
Pred. No.: 0.0039 Length: 1824  
Score: 128.50 Matches: 75  
Percent Similarity: 36.60% Conservative: 52  
Best Local Similarity: 21.61% Mismatches: 129  
Query Match: 4.98% Indels: 91  
DB: 3 Gaps: 12

US-09-830-144-4 (1-504) X US-08-935-855-1 (1-1824)

QY 165 GlyAlaMetAlaValValAlaValLeuLeuAsnAsnLeuLeuValAlaValAlaValGly 184  
DB 736 GGCGCAACAGCGGGTGGCTGTGATCGAGGAGCAAGCACTTATTTGGCCAAATGACAGGA 795  
QY 185 ThrAsnArgAlaLeuLeuCysLysSerThrValAspGlyLeuGlnValThrGlnLeuAsn 204  
DB 796 GACTCTGGCTGTGTGCTGCTCCGAGGCTGGC-----AAAGCTTTAGATATGTCC 843  
QY 205 ValAspHisThrThrGluAsnGluAspGluLeuPheArgLeuSerGlnLeuGlyLeuAsp 224  
DB 844 TATGACCAACAAACAGAGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 897  
QY 225 AlaGlyLysIleLysGlnValGly---IleIleCysGlyGlnGlnSerThrArgArgIle 243  
DB 898 ---GGCAAGGTCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 954  
QY 244 GlyAspTrpLysValLysTrpGlyTrpThrAspIleAspLeuSerAlaIleLysSer 263  
DB 955 GGAGCCACCTCTCAAG-----AGAAAC 978  
QY 264 LysProIleIleAlaGluProGluIleHisGlyAlaGlnPro----- 277  
DB 979 AAAAATCTGCCACCCGAGGAGACAGATATTTCTGCCCTTCTGACATCAAGGTGCTGACT 1038  
QY 278 LeuAspGlyValThrGlyPheLeuValLeuMetSerGlnGlyLeuTrpLysAlaLeuGlu 1095  
DB 1039 CTCACCTGATGACCTGAATTCATGATGATGATGATGATGATGATGATGATGATGATGATG 1095  
QY 298 AlaAlaHisGlyProGlyGlnAlaAsnGlnGluIleAlaAlaMetIleAspThrGluPhe 317  
DB 1096 -----AGAGCCAGAGAGTGTGAGACTTATTCATCAATCAAGATC 1134  
QY 318 AlaLysGlnThrSer-----LeuAspAlaValAlaGlnAlaValAlaVal 331  
DB 1135 AGTCAACGTGATGAAGAAAGGGGAGCTTGGTTATGTGATTCATTCATGGAAGAGCTGCTG 1194  
QY 332 AspArg-----ValLysArgIle-HisSe 339  
DB 1195 GATCAGTGCCTGGGCGCAGACATTCCTGGGATGATGATGATGATGATGATGATGATGATGAT 1254  
QY 339 RasPThrPheAlaSerGly-----G 346  
DB 1255 ATCATCATTTGCTTCAAGCCCGAAACAGATGAGCTTCAAGCAGAGATGCGAAGAGG 1314

QY 346 YGluArgAlaArgPheCysProArgHisGluAspMetThrLeuLeuValArgAsnPheG 366  
DB 1315 AAACGTGAGAGAGGACGTCTTCCACGAGAGGGGCTGAAGACACCGGCAACAGTGAAGAAAG 1374  
QY 366 YTrpProLeuGlyLysMetSerGlnProThrPro-----SerPr 379  
DB 1375 AAGCCAAAGAGGACGTAGTGTCAACCGGACCTTGCCCATGTGACGTGTTTCTGAGCCC 1434  
QY 379 C-AlaPro-----AlaA 383  
DB 1435 TTGACCCGAGACCTGAGCTTTGCTTGTCTTCTTACCTTACCAAGTGGGTATGAGCTGTG 1494  
QY 383 IagGlyArgValTrpProValSerValProTyrSerSerAlaGlnSerThrSerLysT 403  
DB 1495 CAGGGGGCTGGTGTCTTCTTCTCAGC---CAATTACAAAGAGGCCCCCCCCCA 1551  
QY 403 hrSerValThrLeuSerLeuValMetProSerGlnGlyGlnMetValAsnGlyAlaHis 423  
DB 1552 CGCGGACAGCTCGAGGAGCTGCTGCTTAAAGCTTCTTACTCTCTGCGCTCATC 1611  
QY 423 erAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThrLeuThrL 443  
DB 1612 GACTATCGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1671  
QY 443 euGlnSerThrAsnThr 448  
DB 1672 TTTACTCTGTGACACT 1688

RESULT 14  
US-08-935-855-21  
Sequence 21, Application US/08935855  
Patent No. 6066485  
GENERAL INFORMATION:  
APPLICANT: Guthridge, Mark  
TITLE OF INVENTION: NOVEL GROWTH FACTOR INDUCIBLE  
TITLE OF INVENTION: SERINE/THROMBIN PHOSPHATASE, FIN13  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
STREET: Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/935,855  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 1049-1-002 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2081 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Mus musculus



US-09-830-144-4 (1-504) x US-08-258-261B-6 (1-28958)

QY 84 LeuSerAlaGluLeuLeuGlyGlnLeuAsnAlaGluHisAlaGluAlaAspValArg 103  
DB 5506 CTGAAGCCCGCTCTCTCTCTCGGACACCTCATTTGGGACCTGTGGCCGCCACCTCCGC 5565  
QY 104 ArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGlnSerIleAspAsp 123  
DB 5566 GGCCTCTCTT-----TCTCTCCAGAGAC 5586  
QY 124 Ala-----LeuAlaGluIuLysAlaSerLeuGlnSerGlnLeuProGluGlyValPro 140  
DB 5587 GGCTGACCCCTGTGCTGCCCGCCCGGAAGCTCATGCAAGCCCTCCACAGGCGCGCC 5646  
QY 141 GlnHisGlnLeuProGlnTyrGlnLysIleLeuGluArgLeuLysThrLeuGluArg 160  
DB 5647 ATGCTACCCCTCCGAGCTCCGAGGAGAAATGCCGACCTTCCAGCCCTACGAAGGC 5706  
QY 161 GluIleSerGlyValAlaMetAlaValAlaValLeuLeuAsnAsnLysLeuTyrVal 180  
DB 5707 CGAGCTAGCTGCGCC-----CTCATAGGCTCTCTCCAC 5745  
QY 181 AlaAsnValGlyThrAsnArgAlaLeuLeu-----CysLysSerThrValAspGly 197  
DB 5746 GTGCTGCTGCGGATGAAGACCGGTGTGGATGCGCCCGAGGCGGAAGCCCTCGA 5805  
QY 198 LeuGlnValThrGlnLeuAsnValAspHisThrThrGluAsnGlu----- 212  
DB 5806 CGAAGACCAACAGCTGCGCTGCGCCAGCCACCTCCATCCCGCACATGACGGAATG 5865  
QY 213 ---AspGluLeuPheArgLeuSerGln---LeuGlyLeuAspAlaGlyLysIleLysGln 230  
DB 5866 CTGACGACTTCCGCGCTGCGCCAGAGCCACCTACCATCCCGACGATCCCATC 5925  
QY 221 ValGlyIleIleCysGlyGlnGlnSerThrArgArgIleGlyAspTyrLysValLysTyr 250  
DB 5926 ATCTCAACGTACCGCGCGCGCGCCAG-----GACCACAG----- 5964  
QY 251 GlyTyrThrAspIleAspLeuLeuSerAlaAlaLysSerLysProIleIleAlaGluPro 270  
DB 5965 -----CTGCCCTCGCC 5976  
QY 271 GluIle-----HisGlyAlaGlnProLeuAspGlyValThrGlyPhe 284  
DB 5977 GACTACTGGGTCCGCCACTTCCGACACCGCTTCTCTGACGGCGTA----- 6027  
QY 285 LeuValLeuMetSerGluGlyLeuTyrLysAlaLeuGluAlaHisGlyProGlyGln 304  
DB 6028 -----CGTGCCTTACCGCGGAAGG--- 6048  
QY 305 AlaAsnGlnGluIleAlaAlaMetIleAspThrGluPheAlaLysGlnThrSerLeuAsp 324  
DB 6049 -----GCAAGTGTCTTCTCGAGCTCGGGCTCAGCTGTCTCTCC 6090  
QY 325 AlaValAlaGlnAlaValAlaVal-AspArgValLys---ArgIleHisSerAspThrPheAl 343  
DB 6091 GCCCTTGGCAAGACGCTCGAGACAGAGGACGTGCGCATGCGCTTCTCTCC 6150  
QY 343 AserGlyGlyGluArgAla-----ArgPheCysProArgHisGluAspMetTh 359  
DB 6151 ACCCTCCGCAAGGAGGACGAGACGCGCGAGGCTTACCGCGCGC-----TCGGC 6201  
QY 359 rLeuLeuValArgAsnPheGlyTyrProLeuGlyGlyMetSerGlnPro----- 375  
DB 6202 GCTCTCCACTCCGAGCATCACCCGACTGGAGCGCTTCTTCTGCGCCCTTGCCTCCA 6261  
QY 376 -----ThrProSerPro-----AlaProAlaAlaGlyArgValTyrPhe 389  
DB 6262 CGCAAGGTCTCTCTCCACTATGCTTCCAGCGGCGGCTTCTGCGCGGAGCGCTTCC 6321  
QY 389 oValSerValProTyrSerSerAlaGlnSerThrSerLysThrSerValThrLeuSerLe 409  
DB 6322 A-----AGGCACCGCGCGCGGACGTACGACCTTGTCTCG 6357

QY 409 uValMetProSerGlnGlyGlnMetValAsnGlyAlaHisSerAlaSerThrLeuAspG 429  
DB 6358 CTCGAGGGGGGCTCTGTGGCAAGCCATGACGCGGGACCTCGATGGCTCAGCGGTCA 6417  
QY 429 uAlaThrProThrLeuThrAsnGlnSerProThrLeuThrLeuGlnSerThrAsnThrH 449  
DB 6418 CTCACGTGACGGGACGAGCGCGCGCTGCTGCTCTTCCACCTCTCG 6477  
QY 449 sThrGlnSerSerSerSerSerAspGlyGlyLeuPheArgSerArgProAlaHisSe 469  
DB 6478 AGCTTTCGCCAGAGCGGCAAGACAGAGCA-----CGGTGACGCTTGGCGCTAC 6528  
QY 469 rLeuProProGly 473  
DB 6529 CGTATCACTTGA 6541

Search completed: December 10, 2002, 02:07:20  
Job time : 115.518 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus.p2n model

Run on: December 9, 2002, 23:04:29 ; Search time 72.5983 Seconds  
(without alignments)  
2707.033 Million cell updates/sec

Title: US-09-830-144-4  
Perfect score: 2580  
Sequence: 1 MAARRSILQSEQPSWTDD.....AEFFRLMSVDHGQSVVTAP 504

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 350425 seqs, 194966369 residues

Total number of hits satisfying chosen parameters: 700850

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+.p2n.model -DEV=xip  
-Q=/cgn2\_1/USPRO.spool/US09830144/runat\_04122002.141354\_2307/app\_query.fasta\_1.1422  
-DB=Published Applications NA -OPW=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOEXT=0 -UNITS=bites -START=1 -END=1 MATRIX=blonsum62  
-TRANS-human40.cdd -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HRAPIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09830144@cgn\_1.1.51@runat\_04122002.141354\_2307  
-ICPU=6 -ICPU3 -NO XLPXY -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Published Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2580	100.0	1560	9	US-10-158-895-1 Sequence 1, Appl1
2	2580	100.0	1560	12	US-10-123-427-1 Sequence 1, Appl1
3	2580	100.0	1569	9	US-10-158-895-10 Sequence 10, Appl1
4	2577	99.9	1568	9	US-10-158-895-42 Sequence 42, Appl1

5	2575	99.8	1560	12	US-10-123-427-5 Sequence 5, Appl1
6	543.5	21.1	16877	10	US-09-764-877-3349 Sequence 3349, Ap
7	409	15.9	696	10	US-09-925-300-330 Sequence 330, Ap
8	365	14.1	211	10	US-09-864-761-17553 Sequence 17553, A
9	323.5	12.5	455	10	US-09-864-761-769 Sequence 769, Ap
10	271.5	10.5	467	10	US-09-864-761-770 Sequence 770, Ap
11	261	10.1	194	10	US-09-864-761-17554 Sequence 17554, A
12	196.5	7.6	1068	9	US-09-938-842A-2158 Sequence 2158, Ap
13	190.5	7.4	1188	9	US-09-938-842A-958 Sequence 958, Ap
14	188.5	7.3	1074	9	US-09-938-842A-210 Sequence 210, Ap
15	186	7.2	1496	10	US-09-770-445-6 Sequence 6, Appl1
16	178	6.9	1422	9	US-09-935-124A-1 Sequence 1, Appl1
17	175	6.8	1200	9	US-09-938-842A-1634 Sequence 1634, Ap
18	171	6.6	2740	9	US-09-973-941-3 Sequence 3, Appl1
19	171	6.6	2740	10	US-09-973-963-3 Sequence 3, Appl1
20	171	6.6	2740	10	US-09-973-964-3 Sequence 3, Appl1
21	171	6.6	2740	10	US-09-973-077-3 Sequence 3, Appl1
22	171	6.6	2740	10	US-09-973-063-3 Sequence 3, Appl1
23	171	6.6	2740	10	US-09-973-964-3 Sequence 3, Appl1
24	171	6.6	2740	10	US-09-975-072-3 Sequence 3, Appl1
25	171	6.6	2740	10	US-09-972-038-3 Sequence 3, Appl1
26	171	6.6	2740	10	US-09-972-787-3 Sequence 3, Appl1
27	171	6.6	2740	10	US-09-973-965-3 Sequence 3, Appl1
28	169.5	6.6	1272	9	US-09-938-842A-1547 Sequence 1547, Ap
29	168.5	6.5	2099	10	US-09-925-300-715 Sequence 715, Ap
30	167.5	6.5	1305	9	US-09-938-842A-1434 Sequence 1434, Ap
31	164.5	6.4	1059	10	US-09-860-351-3 Sequence 3, Appl1
32	164.5	6.4	1676	10	US-09-860-351-1 Sequence 1, Appl1
33	158	6.1	1143	9	US-09-938-842A-736 Sequence 736, Ap
34	156.5	6.1	1346	10	US-09-828-302-9 Sequence 9, Appl1
35	156.5	6.1	5134	10	US-09-880-107-1608 Sequence 1608, Ap
36	155	6.0	5791	10	US-09-920-300A-1714 Sequence 1714, Ap
37	155	6.0	5791	12	US-10-033-528-1714 Sequence 1714, Ap
38	154.5	6.0	807	9	US-09-938-842A-212 Sequence 212, Ap
39	154.5	6.0	1536	9	US-09-938-842A-1595 Sequence 1595, Ap
40	151.5	5.9	1179	9	US-09-938-842A-1208 Sequence 1208, Ap
41	146	5.7	1014	9	US-09-938-842A-1818 Sequence 1818, Ap
42	145.5	5.6	1410	10	US-09-828-302-10 Sequence 10, Appl1
43	142	5.5	1191	9	US-09-938-842A-1925 Sequence 1925, Ap
44	139.5	5.4	1167	9	US-09-938-842A-1133 Sequence 1133, Ap
45	138.5	5.4	1965	9	US-09-938-842A-1948 Sequence 1948, Ap

## ALIGNMENTS

RESULT 1  
US-10-158-895-1  
Sequence 1, Application US/10158895  
Patent No. US20020155624A1  
GENERAL INFORMATION:  
APPLICANT: ONO, KOICHIRO  
APPLICANT: OHTOMO, TOSHIHIKO  
APPLICANT: TSUCHIYA, MASAYUKI  
TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES  
FILE REFERENCE: 053466/0278  
CURRENT APPLICATION NUMBER: US/10/158, 895  
CURRENT FILING DATE: 2002-06-03  
PRIOR APPLICATION NUMBER: US/09/529, 279  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: PCT/JP98/04796  
PRIOR FILING DATE: 1998-10-22  
PRIOR APPLICATION NUMBER: JP 9/290188  
PRIOR FILING DATE: 1997-10-22  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 1560  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (30)..(1541)  
US-10-158-895-1

## Alignment Scores:

Pred. No.: 2,32e-229 Length: 1560  
Score: 2580.00 Matches: 504  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-830-144-4 (1-504) x US-10-158-895-1 (1-1560)

Qy 1 MetAlaAlaGlnArgSerLeuLeuGlnSerGluGlnGlnProSerTrpThrAspAsp 20  
Db 30 ATGGCGCGCAGAGGAGGAGCTTGTGCGAGAGTGACAGACGACCAAGCTGGACATGAC 89  
Qy 21 LeuProLeuCysHisLeuSerGlyValGlySerAlaSerAsnArgSerTyrSerAlaAsp 40  
Db 90 CTGCTCTCTGCGCACCTCTCTGGGGTGGCTCAGCCCTCCAAACCGCAGCTACTCTGCTGAT 149  
Qy 41 GlyLysGlyThrGluSerHisProGluAspSerTrpLeuLysPheArgSerGluAsn 60  
Db 150 GGCAAGGCACTGAGAGCCACCGCCAGAGGACAGCTGGCTCAAGTTCAGGAGTGAGAAC 209  
Qy 61 AsnCysPheLeuTyrGlyValPheAsnGlyTyrAspGlyAsnArgValThrAsnPheVal 80  
Db 210 AACTGTCTTCTGTATGGGGTCTTCAACGGCTATGATGGCAACCGAGTGACCAACTTCGTG 269  
Qy 81 AlaGlnArgLeuSerAlaGluLeuLeuGlyGlnLeuAsnAlaGluHisAlaGluAla 100  
Db 270 GCCACGGGCTGTCCGAGAGCTCTCTGCTGGGCCAGCTGAATGCCGAGCACGCCGAGGCC 329  
Qy 101 AspValArgArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGluSer 120  
Db 330 GATGTGCGCGCTGTCTGTCTGAGGCGCTTCGATGTGGTGAGAGGAGCTTCTCTGAGTCC 389  
Qy 121 IleAspAspAlaLeuAlaGluLysAlaSerLeuGlnSerGlnLeuProGluGlyValPro 140  
Db 390 ATTGACAGCCCTTGGCTGAGAGGCAAGCCTCCAGTCGCAATTCGACAGGAGTCCCT 449  
Qy 141 GlnHisGlnLeuProGlnTyrGlnLysIleLeuGluArgLeuLysThrLeuGluArg 160  
Db 450 CAGCACCATGCTGCTCTCAGTATCAGAGATCTTTGAGAGACTCAAGACGTTAGAGAG 509  
Qy 161 GluIleSerGlyGlyAlaMetAlaValAlaValLeuLeuAsnAsnLysLeuTyrVal 180  
Db 510 GAAATTTCCGAGGCGGCATGCGCGTGTGGCGTCTTCTCAACCAACAGCTCTACGTC 569  
Qy 181 AlaAsnValGlyThrAsnArgAlaLeuLeuCysLysSerThrValAspGlyLeuGlnVal 200  
Db 570 GCCAATGTCGGTACAAACCGTGCACTTTATGCAATCGACAGTGGTGGTGGAGTG 629  
Qy 201 ThrGlnLeuAsnValAspHisThrThrGluAsnGluAspGluLeuPheArgLeuSerGln 220  
Db 630 ACACAGCTGAACGTGGACCAACACACAGAGAACGAGGATGAGCTCTTCGCTCTTTCCGAG 689  
Qy 221 LeuGlyLeuAspAlaGlyLysIleLysGlnValGlyIleLysCysGlyGlnGluSerThr 240  
Db 690 CTGGCTTGGATGCTGGAAGATCAAGCAGGTGGGATCATCTGTGGCGAGGAGACACC 749  
Qy 241 ArgArgIleGlyAspTyrLysValLysTyrGlyTyrThrAspIleAspLeuLeuSerAla 260  
Db 750 CGCGCGATCGGGATTACAAGTTAAATATGGCTACACGACATTGACCTTCTCAGCGCT 809  
Qy 261 AlalysSerLysProIleIleAlaGluProGluIleHisGlyAlaGlnProLeuAspGly 280  
Db 810 GCCAAGTCCAAACCAATCATCGACAGCCAGAAATCCATGGGCACACGCGCTGGATGGG 869  
Qy 281 ValThrGlyPheLeuValLeuMetSerGluGlyLeuTyrLysAlaLeuGluAlaAlaHis 300  
Db 870 GTGACGGGCTTCTTGGTGTCTGATGCGAGGGGTGTACAAAGGCCCTAGAGGAGGCCCAT 929  
Qy 301 GlyProGlyGlnAlaAsnGlnGluIleAlaAlaMetIleAspThrGluPheAlaLysGln 320

Db 930 GGCCTGGCGAGCCCAACAGGAGATTGTGCGATGATTGACACTGATTGTTGCCAAGCAG 989  
Qy 321 ThrSerLeuAspAlaValAlaGlnAlaValValAspArgValLysArgIleHisSerAsp 340  
Db 990 ACCTCCCTGGACAGTGGCCCGCTGCTGGACCGGGTGAAGCGCATCCACACGCGAC 1049  
Qy 341 ThrPheAlaSerGlyGlyGluArgAlaArgPheCysProArgHisGluAspMetThrLeu 360  
Db 1050 ACCTTCGCCAGTGGTGGGAGCGTGCAGGTTCTGCCCGCGCACGAGGACATGACCTG 1109  
Qy 361 LeuValArgAsnPheGlyTyrProLeuGlyGluMetSerGlnProThrProSerProAla 380  
Db 1110 CTAGTAGGAACTTTGGCTACCCGCTGGGCGAAATAGCCAGGCCACACCGAGCCGAGCC 1169  
Qy 381 ProAlaAlaGlyArgValTyrProValSerValProTyrSerSerAlaGlnSerThr 400  
Db 1170 CCAGCTCGAGGAGACGAGTGTACCTGTGTGTGCCATCTCCAGCGGCCAGAGCACC 1229  
Qy 401 SerLysThrSerValThrLeuSerLeuValMetProSerGlnGlyGlnMetValAsnGly 420  
Db 1230 AGCAAGACACGAGTGACCTCTCCCTGTGTATGCTCCAGCGGCCAGATGGTCAACGGG 1289  
Qy 421 AlaHisSerAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThr 440  
Db 1290 GCTCACAGTGTCTCCACCTGGACGAAGCCACCCCAACCTCCACCAACCAAGCCGAGCC 1349  
Qy 441 LeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerSerSerSerGly 460  
Db 1350 TTAACTCTGAGTCCACCAACACGACGAGCAGCAGCAGCTCCAGCTCTGAGCGAGGC 1409  
Qy 461 LeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGlyArgValGluPro 480  
Db 1410 CTCCTTCGCTCCCGCGCCGCCACTCGCTCCCGCTCCGCGGAGGAGCGTCTGTGTAGCCC 1469  
Qy 481 TyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGlyGluGlnSerVal 500  
Db 1470 TATGTGACATTTGTGAGTTTACCGCTCTGGAGCGTGGACCATGCGAGCAGAGCGGTG 1529  
Qy 501 ValThrAlaPro 504  
Db 1530 GTGACAGCACCG 1541

RESULT 2  
US-10-123-427-1  
; Sequence 1, Application US/10123427  
; Patent No. US20020119525A1  
; GENERAL INFORMATION:  
; APPLICANT: MATSUMOTO, Kunihiro  
; NISHIDA, Eisuke  
; TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/123,427  
; FILING DATE: 17-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/406,854  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: US/08/752,891  
; FILING DATE: 20-NOV-1996  
; APPLICATION NUMBER: JP 8-300856

FILED DATE: 28-OCT-1996  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1560 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1541  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 30..1541  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-123-427-1  
Alignment Scores:  
Pred. No.: 2,32e-229 Length: 1560  
Score: 2580.00 Matches: 504  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
US-09-830-144-4 (1-504) x US-10-123-427-1 (1-1560)  
QY 1 MetaAlaInArgArSerLeuLeuGlnSerGlnGlnProSerTrpThrAspAsp 20  
Db 30 ATGGCGGGCGAGAGAGAGAGCTTCTGCTGAGAGTACAGACAGCCAACTGAGACGATGAC 89  
QY 21 LeuProLeuGlyHisLeuSerGlyValGlySerAlaSerAsnArgSerTrpSerAlaAsp 40  
Db 90 CTGCTCTCTGCTCCACTCTCTGGGGTGGCTCAGCTCCAACTGAGTCTGCTGAT 149  
QY 41 GlyGlyGlyThrGlnSerHisProProGlnAspSerTrpLeuGlyPheAlaGlySerGlnAsn 60  
Db 150 GGCAAGGGCACTGAGAGCCAGCCCGCAGAGGACGCTGCTCAAGTTCAGAGAGGAGAAC 209  
QY 61 AsnGlyPheLeuGlyArgValPheAsnGlyTrpAspGlyAsnArgValThrAsnPheVal 80  
Db 210 AACTGCTTCTGTATGGGGTCTTCAACGGCTATGATGGCAACGAGTACCACTTCTGTG 269  
QY 81 AlaGlnArgLeuSerAlaGlnLeuLeuLeuGlnGlnLeuAsnAlaGlnHisAlaGlnAla 100  
Db 270 GCCACGGCGCTGCTCGCAGAGACTCTGCTGGCCAGCTGATGCGAGACAGCCGAGAGCC 329  
QY 101 AspValAlaArgArgValLeuLeuGlnAlaPheAspValValGlnArgSerPheLeuGlnSer 120  
Db 330 GATGGCGCGCTGTCTGCTGCTCAGGCTTCGATGTGGAGAGAGCTTCTCTGAGTCC 389  
QY 121 IleAspAspAlaLeuAlaGlnValAlaSerLeuGlnSerGlnLeuProGlnGlyValPro 140  
Db 390 ATTAGCGAGCTTGGCTGAGAAAGGCAAGCTCCAGTGCAGATTGCCAGAGGAGTCCCT 449  
QY 141 GlnHisGlnLeuProProGlnTrpGlnLysIleLeuGlnArgLeuGlyThrLeuGlnArg 160  
Db 450 CAGCACCGAGCTCCCTCTCAAGATCAGAAAGATCCTTGAAGACCTCAAGACTTGAAGAGG 509  
QY 161 GlnIleSerGlyGlyAlaMetAlaValAlaValAlaValLeuLeuAsnAsnLysLeuTrpVal 180  
Db 510 GAAATTTGGGAGGGGAGCATGCGCTTGTGGGCGCTCTTCAACAACAAGCTCTACGCTC 569  
QY 181 AlaAsnValGlyThrAsnArgAlaLeuLeuGlySerTrpValAspGlyLeuGlnVal 200

Db 570 GCCATGTGGGTACAAACCGGCACTTTATGCAAAATCGACAGTGGGTCTCAGGTG 629  
QY 201 ThrGlnLeuAsnValAspHisThrThrGlnAsnGlnAspGlnLeuPheArgLeuSerGln 220  
Db 630 ACACAGCTGAACGTGGACCAACACACAGAGAACAGATGAGTCTTCCGCTTCCGACG 689  
QY 221 LeuGlyLeuAspAlaGlyLysIleLysGlnValGlyIleIleIleCysGlyGlnGlnSerThr 240  
Db 690 CTGGCTTGATGTGCTGGAAAGATCAGACAGGTGGGATCATCTGTGGGACAGAGACACC 749  
QY 241 ArgArgIleGlyAspTrpLysValLysTrpGlyTrpThrAspIleAspLeuLeuSerAla 260  
Db 750 CGGCGGATCGGGGATTAACAAGTTAAATATGCTTACACGGCATTTGACCTTCTCAGCGCT 809  
QY 261 AlaySerLysProIleIleAlaGlnProGlnIleHisGlyAlaGlnProLeuAspGly 280  
Db 810 GCCAAGTCCAAACCAATATGACAGAGCCAGAAATCCATGGGACAGCCGCTGGATGG 869  
QY 281 ValThrGlyPheLeuValLeuMetSerGlnGlyLeuTrpLysAlaLeuGlnAlaAlaHis 300  
Db 870 GTGACGGGCTTCTGTCTGCTGATGTGAGGGGTTGTACAAAGCCCTTGAAGAGCCCAT 929  
QY 301 GlyProGlyGlnAlaAsnGlnGlnIleAlaAlaMetIleAspThrGlnPheAlaLysGln 320  
Db 930 GGGCTGGGAGGCGCAACAGAGATGCTGCTGATGATGACCTGATGACCTGTTGCCAAGCAG 989  
QY 321 ThrSerLeuAspAlaValAlaGlnAlaValAlaAspArgValLysArgIleHisSerAsp 340  
Db 990 ACCCTCTTGAAGAGTGGCCCGGCTCTGAGACCGGGTGAAGCGCATCACAGGCAC 1049  
QY 341 ThrPheAlaSerGlyGlyValArgAlaArgPheCysProArgHisGlyAspMetThrLeu 360  
Db 1050 ACCCTGCGCAGTGGTGGGAGGCTCCAGCTTCTCCCGCGGACAGAGACATGACCTTG 1109  
QY 361 LeuValArgAsnPheGlyTrpProLeuGlyGlnMetSerGlnProThrProSerProAla 380  
Db 1110 CTATGAGAGAACTTTGGCTACCCGCTGGCGAAATGAGCCAGCCACCAAGCCAGACC 1169  
QY 381 ProAlaAlaGlyGlyArgValTrpProValSerValProTrpSerSerAlaGlnSerThr 400  
Db 1170 CCAGCTGACAGAGAGAGAGTGAACCTGTGTGTGTCATCTCCAGGCGCCAGAGGCC 1229  
QY 401 SerLysThrSerValThrLeuSerLeuValMetProSerGlnGlyGlnMetValAsnGly 420  
Db 1230 AGCAGACACAGCTGACCTTCTCTGTCATGCTCCAGGCGCAATGTCAGCGG 1289  
QY 421 AlaHisSerAlaSerThrLeuAspGlnAlaThrProThrLeuThrAsnGlnSerProThr 440  
Db 1290 GCTCAAGTCTTCCACCTGAGACAGACACCCCACTCACCACCAACCAAGCCGAC 1349  
QY 441 LeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerAspGlyGly 460  
Db 1350 TTAACTCTCAGTCCACCAACAGCAGACGCGAGACAGCTCCAGCTGAGCGAGGC 1409  
QY 461 LeuPheArgSerArgProAlaHisSerLeuProProGlnGlyLysAspGlyArgValGlnPro 480  
Db 1410 CTTCTCGCTCCCGGCGCGCCACCTGCTCCGCTGGGAGAGAGGTCGTATTAGCCC 1469  
QY 481 TrpValAspPheAlaGlnPheTrpArgLeuTrpSerValAspHisGlyGlnGlnSerVal 500  
Db 1470 TATGTGACTTTGTGATAGTTTACCGCTCTGAGCGTGGACATGGGAGAGAGCGGTG 1529  
QY 501 ValThrAlaPro 504  
Db 1530 GTGACAGCACCG 1541  
RESULT 3  
US-10-158-895-10  
; Sequence 10, Application US/10158895  
; Patent No. US20020155624A1  
; GENERAL INFORMATION:  
; APPLICANT: ONO, KOICHIRO

; APPLICANT: OHTOMO, TOSHIHIKO  
; APPLICANT: TSUCHIYA, MASAYUKI  
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES  
; FILE REFERENCE: 053466/0278  
; CURRENT APPLICATION NUMBER: US/10/158,895  
; PRIOR FILING DATE: 2002-06-03  
; PRIOR APPLICATION NUMBER: US/09/529,279  
; PRIOR FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: PCT/JP98/04796  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: JP 9/290188  
; PRIOR FILING DATE: 1997-10-22  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 1569  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (7)..(1557)  
US-10-158-895-10  
Alignment Scores:  
Pred. No.: 2,34e-229 Length: 1569  
Score: 2580.00 Matches: 504  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-09-830-144-4 (1-504) x US-10-158-895-10 (1-1569)  
QY 1 MetAlaGlnArgSerLeuLeuGlnSerGluGlnProSerTrpThrAspAsp 20  
Db 7 ATGGCGCGCAGAGGAGCTTGTGCGAGGTGAGCAGCAGCAAGCTGGACATGAC 66  
QY 21 LeuProLeuGlyHisLeuSerGlyValGlySerAlaSerAsnArgSerTrpSerAlaAsp 40  
Db 67 CTGCTCTCTGCCACCTCTCTGGGGTGGCTCAGCTTCAACCGCAGCTTCTGCTGAT 126  
QY 41 GlyLeuGlyThrGluSerHisProGluAspSerTrpLeuLysPheArgSerGluAsn 60  
Db 127 GGCAAGGGCACTCAGAGCCACCGCCAGAGGACAGCTGGCTCAAGTTCAAGAGTGAAC 186  
QY 61 AsnCysPheLeuTyrGlyValPheAsnGlyTyrAspGlyAsnArgValThrAsnPheVal 80  
Db 187 AACTGCTTCTGTATGGGTCTTCAACGGCTATGATGGCAACCGAGTGACCAACTTCGTG 246  
QY 81 AlaGlnArgLeuSerAlaGluLeuLeuLeuGlyGlnLeuAsnAlaGluHisAlaGluAla 100  
Db 247 GCCACGGCTGTCCGAGAGCTCTCTGTTGGCCAGCTGAATGCCAGCAGCCGAGGCC 306  
QY 101 AspValArgArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGluSer 120  
Db 307 GATGTGGCGCTGTCTGTGCGAGGCTTCGATGTGTGGAGGAGCTTCTCTGGAGTCC 366  
QY 121 IleAspAlaLeuAlaGluLysAlaSerLeuGlnSerGlnLeuProGluGlyValPro 140  
Db 367 ATTGAGCAGCGCTTGGCTGAGAGGCAAGCCTCCAGTCGCAATTTGCCAGAGGGAGTCCCT 426  
QY 141 GlnHisGlnLeuProProGlnTyrGlnLysIleLeuGluArgLeuLysThrLeuGluArg 160  
Db 427 CAGCACAGCTGCCTCTCTAGTATACAGAGATCTTTGAGAGACTCAAGACGTGAGAGG 486  
QY 161 GluIleSerGlyGlyAlaMetAlaValAlaValAlaLeuLeuAsnAsnLysLeuTyrVal 180  
Db 487 GAAATTTGGGAGGGCCATGGCGTTGTGGCGTCTTCTCAACACAGCTCTACGTC 546  
QY 181 AlaAsnValGlyThrAsnArgAlaLeuLeuCysLysSerThrValAspGlyLeuGlnVal 200  
Db 547 GCCAATGTCGGTACAAACCGTGCACTTTTATGCAATTCAGAGTGGTGGTGGAGG 606

QY 201 ThrGlnLeuAsnValAspHisThrThrGluAsnGluAspGluLeuPheArgLeuSerGln 220  
Db 607 ACACAGCTGAACGTTGGACCAACACAGAGAACGAGGATGAGCTTCCGCTTTTCGAC 666  
QY 221 LeuGlyLeuAspAlaGlyLysIleLysGlnValGlyIleLysCysGlyGlnGluSerThr 240  
Db 667 CTGGCTTGGATGCTGGAAAGATCAACAGCGTGGGATCATCTGTGGCAGGAGGACACC 726  
QY 241 ArgArgIleGlyAspTyrLysValLysTyrGlyTyrThrAspIleAspLeuLeuSerAla 260  
Db 727 CGGCGATCGGGATTAACAAGTTAAATATGCTACACGACATGACCTTCTCAGCGCT 786  
QY 261 AlaLysSerLysProIleIleAlaGluProGluIleHisGlyAlaGlnProLeuAspGly 280  
Db 787 GCCAAGTCCAAACCAATCATCGCAGAGCCAGAAATCCATGGGCGACAGCGCTGGATGG 846  
QY 281 ValThrGlyPheLeuValLeuMetSerGluGlyLeuTyrLysAlaLeuGluAlaHis 300  
Db 847 GTGACGGCTTCTTGGTGTCTGATGTCGAGGGGTGTGTAAGGCCCTTAGAGGACGCCAT 906  
QY 301 GlyProGlyGlnAlaAsnGlnGluIleAlaMetIleAspThrGluPheAlaLysGln 320  
Db 907 GGGCTTGGCAGGCCAACACGAGATGCTGCGATGATTGACATGAGTTTGCCACGAC 966  
QY 321 ThrSerLeuAspAlaValAlaGlnAlaValAspArgValLysArgIleHisSerAsp 340  
Db 967 ACCTCCCTTGGACGCGAGTGGCCAGCGCTGCTGGACCGGTGAAGCGCATCCACAGCGAC 1026  
QY 341 ThrPheAlaSerGlyGlyGluArgAlaArgPheCysProArgHisGluAspMetThrLeu 360  
Db 1027 ACCTTCGCCAGTGTGGGAGCGTGCAGGTTCTGCCCGCCGACAGGACATGACCTGT 1086  
QY 361 LeuValArgAsnPheGlyTyrProLeuGlyGluMetSerGlnProThrProSerProAla 380  
Db 1087 CTAGTAGGAACTTTGGCTACCCCTGGCGAAATGAGCAGGCCACACCGAGCCAGACC 1146  
QY 381 ProAlaAlaGlyArgValTyrProValSerValProTyrSerSerAlaGlnSerThr 400  
Db 1147 CCAGCTGAGAGGAGGAGTACCTCTGTGTGTCATCTACTCCAGCGCCACAGAGCACC 1206  
QY 401 SerLysThrSerValThrLeuSerLeuValMetProSerGlnGlyGlnMetValAsnGly 420  
Db 1207 AGCAAGACCGAGCTGACCTCTCCCTGTCTATGCCCTCCAGGGCCAGATGGTCAACGG 1266  
QY 421 AlaHisSerAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThr 440  
Db 1267 GCTCACAGTGTCTCCACCTGGACGAAGCACCCACCTCCACCAACCAAGCCGAC 1326  
QY 441 LeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerSerSerGlyGly 460  
Db 1327 TTAACCTTGAGTCCCAACACGACACGAGCAGCAGCTCCAGCTCTGACGAGGC 1386  
QY 461 LeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGlyArgValGluPro 480  
Db 1387 CTCTTCGCTCCCGCCCGCCACTCGCTCCCGCTTGGCGAGAGCGTGTGTGAGCCC 1446  
QY 481 TyrValAspPheAlaGluPheTyrArgLeuTyrSerValAspHisGlyGluGlnSerVal 500  
Db 1447 TATGTGACTTTGTGAGTTTATCCGCTCTGGAGCGTGGACCATGGCAGCAGAGCGGT 1506  
QY 501 ValThrAlaPro 504  
Db 1507 GTGACGACCG 1518

## RESULT 4

US-10-158-895-42  
; Sequence 42, Application US/10158895  
; Patent No. US2002015624A1  
; GENERAL INFORMATION:  
; APPLICANT: OHTOMO, KOICHIRO  
; APPLICANT: OHTOMO, TOSHIHIKO  
; APPLICANT: TSUCHIYA, MASAYUKI  
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES

FILE REFERENCE: 053466/0278  
 CURRENT FILING DATE: 2002-06-03  
 PRIOR APPLICATION NUMBER: US/09/529,279  
 PRIOR FILING DATE: 2000-04-11  
 PRIOR APPLICATION NUMBER: PCT/JP98/04796  
 PRIOR FILING DATE: 1998-10-22  
 PRIOR APPLICATION NUMBER: JP 9/290188  
 PRIOR FILING DATE: 1997-10-22  
 NUMBER OF SEQ ID NOS: 48  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 42  
 LENGTH: 1568  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (11)..(1549)  
 US-10-158-895-42

## Alignment Scores:

Pred. No.: 4,43e-229 Length: 1568  
 Score: 2577.00 Matches: 503  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.80% Mismatches: 0  
 Query Match: 99.88% Indels: 0  
 Gaps: 0

US-09-830-144-4 (1-504) x US-10-158-895-42 (1-1568)

Qy 1 MetAlaAlaGlnArgSerLeuGlnSerGluGlnGlnProSerThrThrAsp 20  
 Db 38 ATGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 97  
 Qy 21 LeuProLeuGlnSerGlnSerGlnSerGlnSerGlnSerGlnSerGlnSer 40  
 Db 98 CTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 157  
 Qy 41 GlyIysGlyThrGlnSerHisProProGluAspSerThrLeuLeuPheArgSerGluAsn 60  
 Db 158 GCGAAGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 217  
 Qy 61 AsnGlnPheLeuGlnSerGlnSerGlnSerGlnSerGlnSerGlnSerGlnSer 80  
 Db 218 AACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 277  
 Qy 81 AlaGlnArgSerLeuGlnSerGlnSerGlnSerGlnSerGlnSerGlnSerGlnSer 100  
 Db 278 GCCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 337  
 Qy 101 AspValArgArgValLeuLeuGlnAlaPheAspValValGlnSerGlnSerGlnSer 120  
 Db 338 GATGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 397  
 Qy 121 IleAspAspAlaLeuAlaGlnLysAlaSerLeuGlnSerGlnSerGlnSerGlnSer 140  
 Db 398 ATTGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 457  
 Qy 141 GlnHisGlnLeuProProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160  
 Db 458 CAGCAGCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 517  
 Qy 161 GluIleSerGlnGlnAlaMetAlaValAlaValLeuLeuAsnAsnLysLeuVal 180  
 Db 518 GAAATTTTGGGAGGCGCATGGCGTGTGGCGTCTCTCAACAACAGCTTACGCTC 577  
 Qy 181 AlaAsnValGlyThrAsnArgAlaLeuLeuGlnSerGlnSerGlnSerGlnSerGlnSer 200  
 Db 578 GCCATGTGCGTACAAACCGGCACTTTTATGCAATGCAAGTGGATGGTGGAGGTG 637  
 Qy 201 ThrGlnLeuAsnValAspHisThrThrGluAsnGluAspGluLeuPheArgLeuSerGln 220  
 Db 638 ACACAGCTGAACGTGACCAACACACAGAGAAAGAGATGAGCTTCCGCTTTTCGAG 697

Qy 221 LeuGlyLeuAspAlaGlyLysIleLysGlnValGlyIleLeuCysGlyGlnGlnSerThr 240  
 Db 698 CTGGCTTTGATGCTGGAGAGATCAAGCGAGTGGGATATCTTGGGACAGAGACACC 757  
 Qy 241 ArgArgIleGlyAspThrLysValLysValLysValLysValLysValLysValLysVal 260  
 Db 758 CCGGAGATCGGGATTCACAGGTTAAATATGGCTACACGACATGACCTTCTCAGCGCT 817  
 Qy 261 AlalysSerLysProIleIleAlaGlnProGluIleHisGlyAlaGlnProLeuAspGly 280  
 Db 818 GCAGATCCAAATCAATCATGCAAGCCAGAAATCATGAGGACAGCGCTGATGGG 877  
 Qy 281 ValThrGlyPheLeuValLeuMetSerGlnGlyLeuThrGlyAlaLeuGlnAlaHis 300  
 Db 878 GTACGGGCTTCTTGGTGTGATGCTGAGAGGGGTGTACAAAGCCCTTACAGGACGCCAT 937  
 Qy 301 GlyProGlyGlnAlaAsnGlnGlnIleAlaMetIleAspThrGluPheAlaLysGln 320  
 Db 938 GGGCTGGGAGGCGCAACAGAGATGCTGCAATGATGACATGAGTTGGCCAAAGCAG 997  
 Qy 321 ThrSerLeuAspAlaValAlaGlnAlaValAlaAspArgValLysArgIleHisSerAsp 340  
 Db 998 AACTCCCTGAGCGAGTGGCCAGGCGCTGTGAGACCGGCTGAACGCTCAGAGCGAC 1057  
 Qy 341 ThrPheAlaSerGlyGlyGlnArgAlaArgPheCysProArgHisGluAspMetThrLeu 360  
 Db 1058 ACCTTGGCAGTGTGGGAGCGTGGCAGTGTCTGCCCGGACAGAGACATGACCTTG 1117  
 Qy 361 LeuValArgAsnPheGlyThrProLeuGlyGlnMetSerGlnProThrProSerProAla 380  
 Db 1118 CTAGTGAAGAACTTTGGCTACCCGCTGGCCAAATGAGCACGCCACAGGCCACACC 1177  
 Qy 381 ProAlaAlaGlyIleArgValIleProValSerValProThrSerSerAlaGlnSerThr 400  
 Db 1178 CCAGCTGAGAGAGACAGAGTACCCGTGTGTGTGCTTCTCAGAGGCCACAGAGACC 1237  
 Qy 401 SerLysThrSerValThrLeuSerLeuValMetProSerGlnGlyIleMetValAsnGly 420  
 Db 1238 AGCAAGACCAAGCTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1297  
 Qy 421 AlaHisSerAlaSerThrLeuAspGlnAlaThrProThrLeuThrAsnGlnSerProThr 440  
 Db 1298 GCTCAAGTGTCTTCCACCTGAGAGAGCCACCCACCTTCAACCAAGCCGACCC 1357  
 Qy 441 LeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerSerSerSerSer 460  
 Db 1358 TTAACTCTGAGTCCACCAACAGACACAGACAGACAGACAGACAGACAGACAGACAG 1417  
 Qy 461 LeuPheArgSerArgProAlaHisSerLeuProProGlyGlnAspGlyArgValGlnPro 480  
 Db 1418 CTCTCCGCTCCGCGCCGCGCCACTCGCTCCGCTGCGAGAGACGCTGTGTTAGGCC 1477  
 Qy 481 TyrValAspPheAlaGlyPheThrArgLeuThrSerValAspHisGlyGlnGlnSerVal 500  
 Db 1478 TATGTGACTTGTGCTGATTTTACCGCTCTGAGGCTGGAACATGAGCAGAGAGCTG 1537  
 Qy 501 ValThrAlaPro 504  
 Db 1538 GTACAGACACCG 1549

## RESULT 5

US-10-123-427-5  
 Sequence 5, Application US/10123427  
 Patent No. US20020119525A1

## GENERAL INFORMATION:

APPLICANT: MATSUMOTO, Kunihito  
 NISHIDA, Eisuke  
 TITLE OF INVENTION: TABI PROTEIN AND DNA CODING THEREFOR  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W., Suite 500

CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/123,427  
 FILING DATE: 17-APR-2002  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/09/406,854  
 FILING DATE: <Unknown>  
 APPLICATION NUMBER: US/08/752,891  
 FILING DATE: 20-NOV-1996  
 APPLICATION NUMBER: JP 8-300856  
 FILING DATE: 28-OCT-1996  
 APPLICATION NUMBER: JP 8-126282  
 FILING DATE: 24-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 17981/111  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136

## INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1560 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear

## FEATURE:

NAME/KEY: CDS  
 LOCATION: 30..1541

## FEATURE:

NAME/KEY: mat\_peptide  
 LOCATION: 30..1541  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5:

## US-10-123-427-5

Alignment Scores:  
 Pred. No.: 6,73e-229 Length: 1560  
 Score: 2575.00 Matches: 503  
 Percent Similarity: 99.80% Conservative: 0  
 Best Local Similarity: 99.80% Mismatches: 1  
 Query Match: 99.81% Indels: 0  
 DB: 12 Gaps: 0

## US-09-830-144-4 (1-504) x US-10-123-427-5 (1-1560)

Qy 1 MetAlaAlaGlnArgSerLeuGlnSerGluGlnGlnProSerTrpThrAspAsp 20  
 Db 30 ATGGCGCGCAGAGGAGCTTGGTGCAGAGTGAGCAGCAAGCTGGACAGATGAC 89  
 Qy 21 LeuProLeuGlnSerGlnValGlySerAlaSerAsnArgSerTrpSerAlaAsp 40  
 Db 90 CTGGCTCTCTGCCACCTCTCTGGGGTGGCTCAGGCTCCACCGCAGCTACTCTGCTGAT 149  
 Qy 41 GlyLysGlyThrGluSerHisProProGluAspSerTrpLeuLysPheArgSerGluAsn 60  
 Db 150 GGCAGAGGCACTGAGAGCCACCCCGCAGAGGACAGATGGCTCAAGTTCCAGAGTGAAC 209  
 Qy 61 AsnCysPheLeuValGlyValPheAsnGlyTyAspGlyAsnArgValThrAsnPheVal 80  
 Db 210 AACTGCTCTCTGATGGGGTCTTCAACGGCTATGATGGCAACCGAGTGACCACTTCGTG 269  
 Qy 81 AlaGlnArgLeuSerAlaGluLeuLeuGlyGlnLeuAsnAlaGluHisAlaGluAla 100

Db 270 GCCAGCGGCTGTCCGACAGAGCTCTCTGTGGCGCAGCTGAATGCCGACGACGCCGAGGCC 329  
 Qy 101 AspValArgArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGluSer 120  
 Db 330 GATGTGGCGGCTGTCTGTCTGCTGAGGCCTTCGATGTGGTGAGAGGAGCTTCTTGAGATCC 389  
 Qy 121 IleAspAspAlaLeuAlaGluLysAlaSerLeuGlnSerGlnLeuProGluGlyValPro 140  
 Db 390 ATTGACGACGCTTGGCTGAGAAGCAAGCTCCAGTCGCAATTGCGAGAGGAGTCCCT 449  
 Qy 141 GlnHisGlnLeuProGlnTrpGlnLysIleLeuGluArgLeuLysThrLeuGluArg 160  
 Db 450 CASCACACAGCTGCTCTCTCAGTATCAGAGATCTTCAGAGACTCAAGACGTTAGAGAGG 509  
 Qy 161 GluIleSerGlyGlyAlaMetAlaValAlaValLeuLeuLeuLeuLeuLeuVal 180  
 Db 510 GAAATTTCCGGAGGGGCCATGGCGGTTGTGGCGGCTCTTCTCAACAACAAGCTCTACGTC 569  
 Qy 181 AlaAsnValGlyThrAsnArgAlaLeuLeuLysSerThrValAspGlyLeuGlnVal 200  
 Db 570 GCCAATGTCGGTACAAACCGTGCACCTTTTATGCAATCGACAGTGGATGGGTGCAGGTG 629  
 Qy 201 ThrGlnLeuAsnValAspHisThrThrGlnAsnGluAspGluLeuPheArgLeuSerGln 220  
 Db 630 ACACAGCTGAACGTCGACACACACAGAGAACGAGATGAGCTCTTCGCTCTTCGCGAG 689  
 Qy 221 LeuGlyLeuAspAlaGlyLysIleLysGlnValGlyIleIleCysGlyGlnGluSerThr 240  
 Db 690 CTGGGCTTGGATCTGGAAAGATCAAGCAGGTGGGATCATCTGTGGCAGAGAGACACC 749  
 Qy 241 ArgArgIleGlyAspTrpLysValLysTrpGlyTyThrAspIleAspLeuSerAla 260  
 Db 750 CGCGGATCGGGGATTACAAAGTTAAATATGGCTACACGAGCATTTGACCTTCTCAGCGCT 809  
 Qy 261 AlaLysSerLysProIleIleAlaGluProGluIleHisGlyAlaGlnProLeuAspGly 280  
 Db 810 GCCAAGTCCAAACCAATCATCGACAGCCAGAAATCCATGGGCACAGCGCTGGATGG 869  
 Qy 281 ValThrGlyPheLeuValLeuMetSerGluGlyLeuTyLysAlaLeuAlaHis 300  
 Db 870 GTACCGGCTTCTTGGTGTGATGTCGAGGGGTTGTACAAAGGCCCTTAGAGGAGCCCAT 929  
 Qy 301 GlyProGlyGlnAlaAsnGlnGluIleAlaMetIleAspThrGluPheAlaLysGln 320  
 Db 930 GGCCTGGCGCAGGCCAACCCAGGAGATTGCTCGCATGATTGACACTGATGTTTGCACAGCAG 989  
 Qy 321 ThrSerLeuAspAlaValAlaGlnAlaValValAspArgValLysArgIleHisSerAsp 340  
 Db 990 ACCTCTCTGGACGCGAGTGGCCCGCAGGCGCTGCTGGACCGGGTGAAGCGCATCCACAGCGAC 1049  
 Qy 341 ThrPheAlaSerGlyGlyGluArgAlaArgPheCysProArgHisGluAspMetThrLeu 360  
 Db 1050 ACCTTCGCCAGTGGTGGGAGCGTCCAGGTTCTGCCCCCGCAGCAGGACATGACCTG 1109  
 Qy 361 LeuValArgAsnPheGlyTyProLeuGlyGluMetSerGlnProThrProSerProAla 380  
 Db 1110 CTAGTGAGGAACCTTGGCTTACCCGCTGGCGGAAATGAGCAGGCCACACCGAGCCGAGCC 1169  
 Qy 381 ProAlaAlaGlyGlyArgValTyProValSerValProTySerSerAlaGlnSerThr 400  
 Db 1170 CCAGCTGCAGGAGGAGAGTGTACCTGTGTCTGTGCTATCTCCAGCGCCGACAGACACC 1229  
 Qy 401 SerLysThrSerValThrLeuSerLeuValMetProSerGlnGlyGlnMetValAsnGly 420  
 Db 1230 AGCAAGACGAGCGTGACCTCTCCCTTGTATGCGCTCCCGAGGGCCAGATGGTCAACGGG 1289  
 Qy 421 AlaHisSerAlaSerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThr 440  
 Db 1290 GCTCACAGTGTCTCCACCTGGAGCAAGCACCACCCTCCACCAACCAAGCCGAGCC 1349  
 Qy 441 LeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerSerSerSer 460  
 Db 1350 TTAACTCTGCTGCTCCACCAACGACGACGAGCAGCAGCTCCAGCTCTTGACGGAGCC 1409

Qy 461 LeupheArSerArgProAlaHisSerLeuProGlyGluAspGlyArgValGluPro 480  
Db 1410 CTCTCCGCTCCCGCCCGCCCACTGCTCCGCTGGGAGAGACGGTGTGTGAGCCC 1469  
Qy 461 TyrValAspPheAlaGluPheTyrArgLeuTyrSerValAspHisGlyGluGlnSerVal 500  
Db 1470 TATGTGACCTTGTCTGATGTTTACCGCTCTGGAGCGTGGACCATGGGAGAGAGCGGTG 1529  
Qy 501 ValThrAlaPro 504  
Db 1530 GTGACAGCACCG 1541

## RESULT 6

US-09-764-877-3349  
; Sequence 3349, Application US/09764877  
; Patent No. US20020147140A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005  
; CURRENT APPLICATION NUMBER: US/09/764,877  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3349  
; LENGTH: 16877  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-877-3349

## Alignment Scores:

Pred. No.: 7,18e-40 Length: 16877  
Score: 543.50 Matches: 295  
Percent Similarity: 18.86% Conservative: 43  
Best Local Similarity: 16.46% Mismatches: 107  
Query Match: 21.07% Indels: 1348  
Gaps: 19

US-09-830-144-4 (1-504) x US-09-764-877-3349 (1-16877)

Qy 12 GlnGlnGlnProSerTyrThrAspAspLeuProLeuGlyHisLeuSerGlyValGlySer 31  
Db 1 GAGCGAGCGCCAGCTGACGATGACCTCTCTGCGACCTCTGGGGTGGCTCA 60  
Qy 32 AlaSerAsnArgSerTyrSerAlaAspGlyValGlyHisLeuSerGlyValGlySer 51  
Db 61 GCCTCAACCCCACTACTCTGCTGATGGCAAGGCACTGAGAGCCACCCGCAAGAGAC 120  
Qy 52 SerTyrLeuLysPhe 56  
Db 121 AGCTGGCTCAAGTTCAGGTGTGTGTGCACATTTCTGTTGGCCCGGAGATTGGTT 180  
Qy 56 56  
Db 181 GGTTCGACAGCAAGAAAGACACCACTTGCAGCTTTCTGATGGGCTTGCCAGTGAC 240  
Qy 56 56  
Db 241 ATGTGGCCCTGAGAGGTGCTCTGCTGCTCTTGTGCCAAGGGCTGCTGATGGGTT 300  
Qy 56 56  
Db 301 AGCGTAGCATGGGAGAGAGGGTATCCAGAAATGTCAATAGCCAGAGTAATGATGGCTA 360  
Qy 56 56  
Db 361 AAGCAGGGGAGACCCAGAGGGCCCTGAGAGTGAAGTGTGTGCTTAACTCCGCCA 420  
Qy 56 56  
Db 421 ATTCTTTCCTCTTCTCTCTCACTCTCGTAGACCTGTGTCAGAGGCTCCCTCTGCCC 480

Qy 57 -----ArgSerGluAsnAsnGlyPheLeuTyrGlyValPheAsnGlyTyrAsp 72  
Db 481 TCTCCCTTCCAGAGTGTGAACAACCTGCTCTGTATGGGGTCTTCAACGGCTATGAT 540  
Qy 73 GlyAsnArgValThrAsnPheValAlaGlnArgLeuSerAlaGlyLeuLeuGlnGlyGln 92  
Db 541 GGCACCCAGTGACCACTTGTGGCCAGCGGCTGTCCGAGACTCTGCTGTGGGCTGAG 600  
Qy 93 LeuAsnAlaGluHisAlaGluAlaAspValArgArgValLeuLeuGln----- 108  
Db 601 CTGATGCGCAGACACCGCCGAGCGCGATGTGCGGCTGTCTCTGTGAGGTATGTTGCG 660  
Qy 108 108  
Db 661 GGGCCACAGTAGACCCAGCACCATATGTCGCCACCCAGAGCTTGGCCCTGCACCTC 720  
Qy 108 108  
Db 721 TAGCATGTTGCCAGGGTGTGGTGTGAAGATCTTCGCGCCCTTACCCAGTAGAGAGC 780  
Qy 108 108  
Db 781 AGCTCCAGCGTAGGCCCCCACCACACAGAGTCCAGAGCCAGCAGCAGCAGCAG 840  
Qy 108 108  
Db 841 CCTTAATGCTGGCAGAGAGGTACTGCTGAGACGGGGGATTAGGAGATGGAGCTT 900  
Qy 108 108  
Db 901 GGAGAGAGTGTGAGGTGGAGCAGGCAAGCGCTGTGAATAATGGGTCATTAGAGCT 960  
Qy 108 108  
Db 961 ACCCTTTCTTCTTCTATGTGTGATGCTCAAGCTTCCAGGTGACAGAGCCATCCCTGG 1020  
Qy 108 108  
Db 1021 CGTCCAGAGAGACCTTGCTCCCTTTTGTGAGGGCGCGCCCTCATTTGACTGGTTCC 1080  
Qy 108 108  
Db 1081 ACCTTTGTGCTGGCTGTGGAGATTTCCTCTGAGCTTGTCTTTCAGAGCATGCT 1140  
Qy 108 108  
Db 1141 CAGGCCATTTCAATGAGAAAGATGGGCTTCTGTCGCCGAGAGGGGTGTGCCAGCTT 1200  
Qy 108 108  
Db 1201 TTCTGCTCTTACAGACTCAGCTCAGCTCATTGCAAGTATTTTCAGCAGATTCACACAG 1260  
Qy 108 108  
Db 1261 GGGAGAAGTGTCACTGTACCAAGTCTCTGGGGCTGTGGGGTTTGAAGAGAGCTGCC 1320  
Qy 108 108  
Db 1321 AGAAAGAGTTAGTATATCATACAGAGCTGTATAGCCTTGAGCTTTCTGAGCAGTT 1380  
Qy 108 108  
Db 1381 GCCTTGCTGAGTGCAGGTTGGACTGTCTTAACTTAACCCCTGGAAACAACATTAACTCC 1440  
Qy 108 108  
Db 1441 GTTGTGTGAGACTGAGGGGGCCAGAGGTCAACACAGCTGGGCTTACGCCAAGCTTGTCT 1500  
Qy 108 108  
Db 1501 CTTAGCTTCCGACTGTCTCTGCTCCAGCCGCTCTGACAGTGTGCTCTGAGAGGTGC 1560



Qy 108 ----- 108  
Db 1561 ACACCCCTTCGAGGCTGAAGGCTTTGTCAAAGACATTGATCTCGAGGAAGCAGCCGGTG 1620  
Qy 108 ----- 108  
Db 1621 CCTTGCA GTGCTGGGCCAGAGGCAAGACTGACATGTGGAAAGCTCCATCACACAAGAAC 1680  
Qy 108 ----- 108  
Db 1681 TGCAGTGAAGACAGCAAGCTGCTGCTGTGATTAATAGAGGACATTTGGCACCAAGTGAC 1740  
Qy 109 ----- AlaPheAspValVal 113  
Db 1741 AGTGGTGTTTGAACACAGCCTTTGGCCCTGCTGTGTGTCCTCCCTAGGCTTCGATGTGGTG 1800  
Qy 114 GluArgSerPheLeuGluSerIleAspAlaLeuAlaGluLysAlaSerLeuGlnSer 133  
Db 1801 GAGAGGAGCTTCCTGGAGTCCATTGACGACGCTTGGCTGAGAAAGCAAGCCTCCAGTCG 1860  
Qy 134 GlnLeuProGlu----- 137  
Db 1861 CAATTGCCAGA-GGTAAATTTCCCGAGCCGACACCCAGGGAGTCAAGTCCAGGCCAGCT 1919  
Qy 137 ----- 137  
Db 1920 TTGCAAGGAGCATGGACTCATCTACTTTCTTGACATTACTGGCCAGAGCAACAGGCGTT 1979  
Qy 137 ----- 137  
Db 1980 AGGGAGCAGTTCTGTATGGTGACACTGGTGTGGGCCACAGGTGAGGAGCTCGCTGCT 2039  
Qy 137 ----- 137  
Db 2040 CTGCTGTGGCAGCTGTCATGCTAAAGCTCAGCAGTGTCTGGGCAGTGGTGACATGAGGA 2099  
Qy 137 ----- 137  
Db 2100 AGGGCTTTTCCAGTCTTCGTCGTAATGGTCCCGCCCTTGTCAACCACTGCTCAC 2159  
Qy 137 ----- 137  
Db 2160 ATCGCTACTGCCACCTGCAGCGCTTACCTCTGTGCTGTGAGTTGCGTCTATCCAGCTT 2219  
Qy 137 ----- 137  
Db 2220 TCTGGAAGAACCTTAGCCTGGAGCTGTAGCGAAGGCTTCATCTGGCTGTGCCATGAGCC 2279  
Qy 137 ----- 137  
Db 2280 ATCTGTCTGTCTCCCCGGGCCCTCGGATCTTTTGCTCTTAAGAGCAAGGAGGCGCG 2339  
Qy 137 ----- 137  
Db 2340 GCATGTGGCTCAGGCTGTAAATCCAGCACCTTGGGAGGCCGAGGCGCGGATCACAA 2399  
Qy 137 ----- 137  
Db 2400 GGTCAAGAGATCAAGACCATCTGGCTAACATGGTGAACCCCGTCTCTACTAAAAATAC 2459  
Qy 137 ----- 137  
Db 2460 AAAAAATTAGCTGGGGTGGTGGGGCGCCTTGTAGTCCAGCTACTCTCGAGGCTGAA 2519  
Qy 137 ----- 137  
Db 2520 GCAGGAGATGTGTGAACCTGGGAGGCGAGCTTGTAGTGAGCCGAGATCATGCCACTG 2579  
Qy 137 ----- 137  
Db 2580 CACTGTAGCCTGGGGGACACAGCGAGCTCCATCTCAAAAAAAAAAAAAAGAGTAAGGG 2639  
Qy 137 ----- 137

Db 2640 AGTGGAGAGGGCTAGGAAGATGGTCCAGAGTTCTGTCTCTGCCCTGACCCCTCTGTGATGG 2699  
Qy 138 ----- GlyValProGlnHisGlnLeuProGlnIleLeuGluArgLeu 155  
Db 2700 TTGTAGGGAGTCCCTCAGCACCAGCTGCTCTCTCAGTATCAGAAGATCCTTTGAGAGACTC 2759  
Qy 156 LysThrLeuGluArgGluIleSerGlyGlyAlaMetAlaValValAlaValLeuLeuAsn 175  
Db 2760 AAGACGTTAGAGGGGAAATTTCCGGAGGGCCATGCGCGTGTGGCGCTCTTCTCAAC 2819  
Qy 176 AsnLysLeuTyxValAlaAsnVal----- 183  
Db 2820 AACAACTCTACGTCCGCAATGTCCGTGAGCCCTCTCTGCCAGGCGAGGAGACTG 2879  
Qy 183 ----- 183  
Db 2880 GGGAGAGGTCAAGCCACAGGGGTGGTGCAITTAATTTGCAATCTCTTTCCAGACACTTCA 2939  
Qy 183 ----- 183  
Db 2940 CGCACTTTAAACCCAGGCTCTCTGAGACCGTTGGGTATGTCCCTCTCCACAGTGAGCC 2999  
Qy 183 ----- 183  
Db 3000 TCAGTCCCAACTGGAAGGAGAAAGACGGATGGGAGACAGGTGTCTGGCCCTTTAGTC 3059  
Qy 183 ----- 183  
Db 3060 CCTATTCTCTTTAACTCATCTCTGGACAGTCTATCCCATCTGGGTGGCTTTGGGG 3119  
Qy 183 ----- 183  
Db 3120 GCGCGTCTTCTGGAAGAGGTTTACCTGGCCATGAATAACCAACAGGATCCCATGAACA 3179  
Qy 183 ----- 183  
Db 3180 GCTGCAGATGGGCTAGGTGACAGGACATTTGGGTTTGTGAGAACACAGACAGAGTGC 3239  
Qy 183 ----- 183  
Db 3240 AGGTTTCAAGTAAAGGACTCTGTAGAGACCTTCTGTATGATGCTGCCCTTTTAAATAC 3299  
Qy 183 ----- 183  
Db 3300 TCCAGTGGAAATTTTCCACAATATAGAACATAGAGTGAATTAATGAACATTC 3359  
Qy 183 ----- 183  
Db 3360 ATGCGCCATCCCAATCCAGAGTTTATCAACTGTGGCCAGCTCTTTGTGCCACGCT 3419  
Qy 183 ----- 183  
Db 3420 TCTCTATTTGAAGCAATGCCAGACACTGTATCATGTTATCTGTAAATATTTCACTCTG 3479  
Qy 183 ----- 183  
Db 3480 TATCTCTAGAGAAGAGTCTGTTTATTAAGATCATTAATCCTATGATTACACTGAAAAAG 3539  
Qy 183 ----- 183  
Db 3540 TTAACCCCAATTTTTCGTATCACAGCTGTCCCTAACTAGTGACTCACCCCGTTGGTCT 3599  
Qy 183 ----- 183  
Db 3600 GAGCCTGTTTGGCCATTTCCAGTATTTCCATGTGTGAATGCTGCTCTTCCCTCTCT 3659  
Qy 183 ----- 183  
Db 3660 GCCTTCCCGTATGCCCTATTCTCTCTGTGTGTAGTCTTTGTGCTAGCTGTTCCACTTCT 3719  
Qy 184 ----- GlyThrAsnArgAlaLeuLeuCysLysSerThrValAspGlyLeuGlnValThr 201  
|||||



```

Db 3720 GCCACAGGTACAAACCGTGCACCTTTATGCAATGACAGTGGATGGTGACAGGTGACA 3779
Qy 202 GlnLeuAsnValAspHisThrThrGluuAsnGluuAspGluLeuPheAspLeuSerGluLeu 221
Db 3780 CAGCTGAACGTGGACCAACACACAGAGAAAGAGATGAGTCTCTCCGCTTCTTCGACGCTG 3839
Qy 221 -----
Db 3840 GGTGAGTGGGAGAGTGGAGCGGAGAGCTGATCCCATGGGCTCACCTTCGCTGCTT 3899
Qy 221 -----
Db 3900 TGTGTGTGGGTAGAGAGCGGTGTGTAGAGGGCTGTGATCTTGAGCTCCCAAGCAGC 3959
Qy 221 -----
Db 3960 CTGCTGGGGTTCATTCGCCAGACCTGCCGCTTACTGTGTGTGAATTGAAGTTT 4019
Qy 221 -----
Db 4020 CTGTGCTCAGATCCCCGCTGTGTAAAGATTAAGTAACCTAAGCTCGGGATGTAG 4079
Qy 221 -----
Db 4080 GAGGATTCAGTTAGTGCATGTGAATGCTTCTGAGAGTCCGCGACACAGAAAGACACTT 4139
Qy 221 -----
Db 4140 ACTGTTATGTGTGAGACTGAAAGAGGCCAAAGAGTCCAGAGAGCCAGCTGCTGCTGA 4199
Qy 221 -----
Db 4200 GCTGCTCCTCCCTCCCAAGTGAAGCTCAAGGACAGATGTGGGGACACAGGGATGGGACAGG 4259
Qy 221 -----
Db 4260 AGACGGCAGCAGAGCTGCTCCGTCAGGTGTGCTGAGAGACGCGGCTGTGATGAGGT 4319
Qy 221 -----
Db 4320 GTGTGTATGAGGGCTGGGGACCTAGAGACACAGGAGCTTTGGGTAGCTGCTCCAGC 4379
Qy 221 -----
Db 4380 GTGCCCTGTGTGTCTCAATTCATTCAGTGGGTCTTATTGCTTCTTCCATGAC 4439
Qy 221 -----
Db 4440 TGTGTCTGTCCCTCTTTTGTCTTGTGTGAACAAGAGAGATTTGCACT 4499
Qy 222 -----
Db 4500 GTTTCCTCCGTAGGCTTGTGATGCTGGAAGATCAAGAGTGGGATCATCTGTGGGA 4559
Qy 223 ngluSerThrAspValIleGluAspValIleValValValValValValValValValVal 257
Db 4560 GGAGAGCAACCGGAGGATTCAGAGGTTAAATATGCTTAACAGCAATTTGACCT 4619
Qy 257 uLeuSer-----AlaAlaValSerLysProIleIleAlaGluProGluIleHisGlyAl 275
Db 4620 TCTCAGTAGGTGCGCCAGCGCTGCTGCTTGAAGAACAGAA----- 4668
Qy 275 aglnProLeuAspGlyValThrGlyPheLeuValLeuMetSerGluGly-LeuTyrLysAs 295
Db 4669 -GGTCTTGGGGAGGCCAAGATGGAGATTAATGTGCTG-----GGAGCTCAGGGCTG 4721
Qy 295 lAlaGluAlaAlaHisGly-----ProGlyGlnAlaAsnGlnGluIleAla- 310
Db 4722 CAGTAGCCGTGATCATCCACTGCATCCAGCTCGGGCAACAGAGAACCCGTGTCTC 4781
Qy 311 -----AlaMetIleAspThrGluPheAlaLysGlnThrSerLeuAspAlaValAla 328
Db 4782 AAAAAAAAAAAAAAGGTCCAGAAAGCTGGGGCAGAGCC---AGACTGTAGGCCCCAGTGA 4838

```

```

Qy 328 lAlaValAlaAspArgValIleArgIleHisSerAspThrPheAlaSerGlyGlyAla 348
Db 4839 CACATGTGCAGTCAAAATTAATTACAGAGCCCAAGCTTCTTCTCC---CCAGAGGA 4895
Qy 348 rg-----AlaArgPheCysProArgHisGluAspMetThrLeuValArgAsn 365
Db 4896 GGTGTGTTTGGGCGAGGTGAGCCAGAGGAGGAGTGTGGTACTGTAGAAACAAAGG 4955
Qy 365 heGlyTyrProLeuGlyGluMet-----SerGlnProThrPro-----S 378
Db 4956 GTGAGTTGAAGCTGGCCAAAGCTGAGGGCAGCAGCTAGTGCACAAAGAGGAGCGGA 5015
Qy 378 erProAlaProAlaAlaGlyIleArgValTyr-ProValSer-----ValProTyrSer 396
Db 5016 AGCCAGAGTGGCATGTGATGAGACAGACAGCGCTGTGTGCTGCCATGGCTGC 5075
Qy 396 erAlaGlnSerThrSerLysThrSerValThrLeuSerLeuValMetProSerGln---- 414
Db 5076 AGCAGAGG-----TGTGAGATGAATTTGTTATTTCCCTCTGAGACA 5117
Qy 415 -----GlyGlnMetValAsnGlyAlaHisSerA 424
Db 5118 CCTCTCTGACATCACCCAGCGCTCAGTGCAGCAGAAATGGCGAGTCATCATCACAGG 5177
Qy 424 lAserThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThrLeuThrLeuG 444
Db 5178 CCAGCCCGTGCAGAGCTGTGATGTCCTTTTATGATTCATTCCTACTCAGGCCCT 5237
Qy 444 lAserThrAsnThrHis-ThrGlnSerSerSerSerSerSerSerSerSerSerSerSer 463
Db 5238 GAGGAGCATCTGCCATCACCCAGAGGCATTTTTCATAGAGGAAACAGAGCATGAGAGA 5297
Qy 464 SerArgProAlaHisSerLeuProProGly 473
Db 5298 CATCAGCAGCCAC-----CCTGGT 5318

RESULT 7
US-09-925-300-330
; Sequence 330, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 330
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (643)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (657)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (685)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-330

Alignment Scores:
Pred. No.: 2,31e-29 Length: 696
Score: 409.00 Matches: 78

```

Percent Similarity: 98.75% Conservative: 1  
Best Local Similarity: 97.50% Mismatches: 1  
Query Match: 15.85% Indels: 0  
DB: 10 Gaps: 0

US-09-830-144-4 (1-504) x US-09-925-300-330 (1-696)

Qy 425 SerThrLeuAspGluAlaThrProThrLeuThrAsnGlnSerProThrLeuThrLeuGln 444  
Dy 13 GCCACCTCGACGAGCCACCCACCTTACCAACCAAGCCGACCTTAACCTCGAG 72  
Qy 445 SerThrAsnThrHisThrGlnSerSerSerSerSerSerSerSerSerSerSerSer 464  
Dy 73 TCCACCAACACGACACGACGAGCAGCAGCTCCAGCTCTACGAGGAGGCTCTCCGCTCC 132  
Qy 465 ArgProLahisSerLeuProProGlyGluAspGlyArgValGluProTyrValAspPhe 484  
Dy 133 CGGCCCCCCTCGCTCCGCTCGGAGGACGGTCTGTGTGAGCCCTATGTGGACTTT 192  
Qy 485 AlaGluPheTyrArgLeuTrpSerValAspHisGlyGluGlnSerValValThrAlaPro 504  
Dy 193 GCTGAGTTTACCGCTCTGGAGCGTGGACCATGCGGACGAGCGTGGTGACAGCACCG 252

## RESULT 8

US-09-864-761-17553  
; Sequence 17553, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeonica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 17553  
; LENGTH: 211  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO Z83845.14  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.77  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.73  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.78  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4  
; OTHER INFORMATION: NT HIT: gi5174702, EVALUE 1.00e-115  
; OTHER INFORMATION: SWISSPROT HIT: Q15750, EVALUE 5.00e-36  
; OTHER INFORMATION: EST\_HUMAN HIT: BE988567.1, EVALUE 1.00e-115  
US-09-864-761-17553

## Alignment Scores:

Pred. No.: 5.14e-26 Length: 211  
Score: 365.00 Matches: 70  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 14.15% Indels: 0  
DB: 10 Gaps: 0

US-09-830-144-4 (1-504) x US-09-864-761-17553 (1-211)

Qy 312 MetIleAspThrGluPheAlaLysGlnThrSerLeuAspAlaValAlaGlnAlaValVal 331  
Dy 1 ATGATTGACACTGAGTTTCCCAAGCAGACCTCCCTGGACGACGTCGCCAGGCGCTCGTG 60  
Qy 332 AspArgValLysArgIleHisSerAspThrPheAlaSerGlyGlyGluArgAlaArgPhe 351  
Dy 61 GACCGGTGAAGCGCATCCACAGCGACACCTTCGCCAGTGGTGGGAGCGTCCAGGTTTC 120  
Qy 352 CysProArgHisGluAspMetThrLeuLeuValArgAsnPheGlyTyrProLeuGlyGlu 371  
Dy 121 TGGCCCCGCGCAGGACATGACCTGCTAGTGAGGAACCTTTGGCTACCCGCTGGGCGAA 180  
Qy 372 MetSerGlnProThrProSerProAlaPro 381  
Dy 181 ATGAGCCAGCCACACCGAGCCAGCCCA 210

## RESULT 9

US-09-864-761-769  
; Sequence 769, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeonica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27

```

; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 769
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO 283845.14
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.77
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.73
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.78
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.4
; US-09-864-761-769
Alignment Scores:
Pred. No.: 1,01e-21 length: 455
Score: 323.50 Matches: 73
Percent Similarity: 79.00% Conservative: 6
Best Local Similarity: 73.00% Mismatches: 12
Query Match: 12.54% Indels: 9
DB: 10 Gaps: 3
US-09-830-144-4 (1-504) x US-09-864-761-769 (1-455)
OY 274 G|y|a|a|g|n|p|o|l|e|u|a|p|g|y|a|t|t|g|l|p|h|e|u|v|a|l|e|u|m|e|t|s|e|r|g|l|u|g|l|e|u|t|r 293
Db 168 G|G|C|G|T|T|G|A|C|C|A|C|T|G|A|T|-----CTC|C|T|G|A|T|T|T|A|G|C|T|C|C|A|G|A|T|T|A|G|G|C 215
OY 294 Lys|Ala|Leu|Gln|Ala|Ala|His|Gly|-----Pro|Gly|Gln|Ala|-----Aen|Gln|Iu 308
Db 216 CAT|G|C|C|C|C|C|C|C|G|G|C|C|G|T|C|C|T|T|A|C|A|G|G|T|C|T|C|C|A|C|C|C|C|C|C|C|G|G|A|G 275
OY 309 I|l|e|a|a|a|a|e|t|t|l|e|a|s|p|t|t|g|l|p|h|e|a|l|a|y|s|c|l|n|t|t|s|e|r|e|u|a|p|a|l|a|a|a|g|n 328
Db 276 A|T|T|G|C|T|G|G|A|G|A|T|T|G|A|C|A|G|A|T|T|T|G|C|C|A|G|A|G|A|C|C|T|C|C|G|A|G|G|A|G|G|C|C|A|G 335
OY 329 A|A|a|a|V|a|a|s|P|a|r|q|V|a|l|l|y|e|a|r|g|l|e|H|I|s|e|r|s|p|t|t|p|h|e|a|l|a|s|e|r|g|l|y|g|l|u|a|r|g 348
Db 336 G|C|G|C|G|T|G|G|A|C|C|G|G|T|A|A|G|C|G|A|T|C|C|A|G|C|A|G|C|A|C|C|T|T|G|C|G|A|G|T|G|G|G|A|G|C|C|T 395
OY 349 A|A|a|t|g|p|t|e|C|y|s|p|t|o|a|r|g|H|I|s|G|l|u|a|s|p|e|t|t|t|e|u|l|e|u|V|a|l|a|r|g|a|n|p|t|e|g|l|Y|T|r|P|ro 368
```

```

Db 396 G|C|C|G|G|T|T|G|C|C|C|G|C|G|A|C|A|G|A|C|A|T|G|A|C|C|T|G|T|A|G|A|G|A|A|C|T|T|G|G|C|T|A|C|C|G 455
RESULT 10
US-09-864-761-770
; Sequence 770, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 770
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO 283845.14
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 5.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
; US-09-864-761-770
```

Alignment Scores: 6.64e-17 Length: 467  
Pred. No.: 271.50 Matches: 72  
Score: 56.74% Conservative: 8  
Percent Similarity: 51.06% Mismatches: 23  
Best Local Similarity: 10.52% Indels: 38  
Query Match: 10 Gaps: 5  
DB:

US-09-830-144-4 (1-504) x US-09-864-761-770 (1-467)

QY 1 MetAlaGlnArgSerLeuGlnSerGluGlnProSer----- 16  
Db 56 GTGGCCCGTGAGAGGTGGCTCTGCTGCTTGTCCCAAGGGCTGCTCTGATGGGGTAG 115  
QY 17 -----TTPThrAspAspLeuProLeuGlnSerGlyValGlySerAlaSer 33  
Db 116 CGTGACATGGGAGAGGTATCCCAATGTCTAT-----ACCCAGAT 160  
QY 34 AsnArgSerTyrSerAlaAspGly-----LysGlyThrGluSer----- 46  
Db 161 GAAATGATGGCTAAAGCAGGGGGACCCAGAGGGCCCTGAAGCTGCAGCTGTGCTGCT 220  
QY 47 -----HisProGluAspSerTrpLeuLys 55  
Db 221 TTAGTCTCCCAATCTCTTCCCTTCTCCCTCCACTCCGTGAGACCCCTGCTCTCAGG 280  
QY 56 -----Phe-ArgSerGluAsnAsnAspPheLeuTyrGlyValPh 68  
Db 281 CTTCCCTCTGCTCCCTCTCTCCAGAGTGAGAACAACTGCTCTGATGGGGTCTT 340  
QY 68 eAnGlyTyrAspGlyAsnArgValThrAsnPheValAlaGlnArgLeuSerAlaGluLe 88  
Db 341 CAACGGCTATGATGACCAACCGAGTGACCACTTCGTGGCCAGGGCTGTCACAGGT 400  
QY 88 uLeuLeuGlyGlnLeuAsnAlaGluHisAlaGluAlaAspValArgValLeuLeuGln 108  
Db 401 CTTGCTGGGCCAGCTGAATGCCAGACGCCAGCCGCGATGTGGGGCTGTGCTGTGCA 460  
QY 108 n 108  
Db 461 G 461

RESULT 11  
US-09-864-761-17554  
; Sequence 17554, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wengsheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aetomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 17554  
; LENGTH: 194  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO 283845.14  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 8  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10  
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 5.3  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.7  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.7  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.4  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9  
; OTHER INFORMATION: NT HIT: g111418147, EVALUE 3.00e-82  
; OTHER INFORMATION: SWISSPROT HIT: Q15750, EVALUE 8.00e-24  
; OTHER INFORMATION: EST\_HUMAN HIT: AL118967.1, EVALUE 5.00e-82  
US-09-864-761-17554

Alignment Scores:  
Pred. No.: 1.84e-16 Length: 194  
Score: 261.00 Matches: 51  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.12% Indels: 0  
DB: 10 Gaps: 0

US-09-830-144-4 (1-504) x US-09-864-761-17554 (1-194)

QY 58 SerGluAsnAspCysPheLeuTyrGlyValPheAsnGlyTyrAspGlyAsnArgValThr 77  
Db 2 AGTGAGAACAACTGCTTCTGTATGGGGTCTTCAACGGCTATGATGGCAACCGAGTGACC 61  
QY 78 AsnPheValAlaGlnArgLeuSerAlaGluLeuLeuGlyGlnLeuAsnAlaGluHis 97  
Db 62 AACTTCGTGGCCAGCGCTGCTCCGACAGCTCTGCTGGCCAGCTGAATGCCGAGCAC 121  
QY 98 AlaGluAlaAspValArgValLeuLeuGln 108  
Db 122 GCCGAGGCCGATGTCGGCGTGTCTGCTGCGAG 154

RESULT 12  
US-09-938-842A-2158  
; Sequence 2158, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong



```

Db 226 TTTGTTGGCGTCTACAGCGGCACGGTGGCTGAGACATCTCGCTTCATCATGATCAT 285
QY 84 LeuSerAlaGluLeuLeuGlyGlnLeuAsnAlaGluHisAlaGluAlaAspValArg 103
Db 286 ATGTTCCACCAC-----CTAAGAGGTTTACTGCAGACAA---CAGTGTATGTCTATCA 336
QY 104 ArgValLeuLeuGlnAlaPheAspValValGluArgSerPheLeuGluSerIleAsp 123
Db 337 GAGGTGATAAAGGCGTTCAGCCACTCAAGAGCGTTCATCCATGATGTTACAAAT 396
QY 124 AlaLeuAlaGluLeuAlaSerLeuGlnSerGlnLeuProGluGlyValProGlnHisGln 143
Db 397 -----CAATTTCAA 405
QY 144 LeuProProGlnTyrGlnLysIleLeuGluArgLeuLysThrLeuGluArgGluIleSer 163
Db 406 ACTAGACCTCAGATAGCCACAGTG----- 429
QY 164 GlyGlyAlaMetAlaValAlaValLeuLeuAsnLysLeuTyrValAlaAsnVal 183
Db 430 ---GGATCATGCTGCTTGAAGTGCATCTCGATGGGAAGCTATACGTGGCCACGCA 486
QY 184 GlyThrAsnArgAlaLeuLeuCysLys---SerThrValAspGly---LeuGlnValThr 201
Db 487 GGGGACTCACGGCGCGTCTGGGCAAGTCATGAGGGTAAACAGGTGAAGCTCATGCCACT 546
QY 202 GlnLeuAsnValAspHisThrThrGluAsnGluAspGluLeuPheArgLeuSerGlnLeu 221
Db 547 CAGCTCTCAGCAGACGACCAACGCATCTATAGATCAGTCAGTCAGACGGGAATTCAGGCCCTG 606
QY 222 GlyLeuAspAlaGlyLysIle-----LysGlnValGlyIle 233
Db 607 CATCCGGATCATCCAGATATTGTGTTCTGAAACATAACCTCGCGGAGTAAAGGCATC 666
QY 234 IleCysGlyGlnGluSerThrArgArgIleGlyAspTyrLysValLysTyrGly----- 251
Db 667 ATT-----CAGTTTCAGATCCATCTGGTGATGTGATTGTAAGAGGTCAAGTTC 717
QY 252 -----TyrThrAspIleAspLeuLeuSerAlaLysSerIle 266
Db 718 AACAGGGAACCACTATATGCAAAATTCGGCTGAGGTCAACCGTTC---AGCAAGCCATTA 774
QY 267 Ile---AlaGluProGluIle-----HisGlyAlaGlnProLeuAspGlyValThrGly 283
Db 775 CTGAGTCAGACCGCGCATCACGGTCATACACTGAGCGCGCACGAT-----CAG 825
QY 284 PheLeuValLeuMetSerGlyLeuTyrLysAlaLeuGluAlaHisGlyProGly 303
Db 826 TTCATTATATGTGCATCAGATGGACTGTGGCAACATATG----- 864
QY 304 GlnAlaAsnGlnGluIleAlaAlaMetIleAspThrGlu-----PheAlaLys 319
Db 865 ---AGCAACCAAGAGCAGTACATAGTCAGATCATCCGGAACCGGGATAGCAAG 921
QY 320 GlnThrSerLeuAspAlaValAlaGlnAlaValValAspArgValLysArgIleHisSer 339
Db 922 CGGTGGTGAAGTAGCGCTCAAGAAGCGCAAGAGAGAGAGATGAGATCTCAGAC 981
QY 340 AspThrPheAlaSerGlyGlyGluArgAlaArgPheCysProArgHisGluAspMetThr 359
Db 982 CTGAAAAAGATAGACAGAGGAGTAGGAGACATTTC-----CACGATGACATAACA 1032
QY 360 LeuLeuValArgAsnPheGlyTyrProLeuGlyGluMetSerGlnProThrProSerPro 379
Db 1033 GTGATTGTGTTCTCTTTGTACAAACCTA-----GTGAGCAGAGGAAGATGTTGAGA 1086
QY 380 AlaProAlaAlaGlyArgValTyrProValSerValProTyrSerSer---AlaGln 398
Db 1087 GCACCAGCGGTGTACGTAGGAGGAGCGGGTGTGAATCTACCTCACAAACCCCTGGCGCT 1146
QY 399 SerThrSerLysThr 403
Db 1147 TGCACACGCCCCACT 1161

RESULT 14
US-09-938-842A-210
; Sequence 210, Application US/099388842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krebs, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE OF INVENTION: SAME, AND METHODS OF USE
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 210
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-210

Alignment Scores:
Pred. No.: 9,69e-09 Length: 1074
Score: 188.50 Matches: 90
Percent Similarity: 44.86% Conservative: 67
Best Local Similarity: 25.71% Mismatches: 135
Query Match: 7.31% Indels: 59
DB: 9 Gaps: 16

US-09-830-144-4 (1-504) x US-09-938-842A-210 (1-1074)
QY 59 GluAsnAsnCysPheLeuTyrGlyValPheAsnGlyTyrAspGlyAsnArgValThrAsn 78
Db 142 GATAACACTTCCTCTTCG---GGTGTCTATGATGGTAAAGTGTTCCTAAG 198
QY 79 PheValAlaGlnArgLeuSerAlaGluLeuLeuLeuGlyGlnLeuAsnAlaGluHisAla 98
Db 199 TTTGTGCCAAGTATCTACACGACGAGTCTTCTAGTGATGAG-----CGGTATGCA 249
QY 99 GluAlaAspValArgValLeuLeuGlnAlaPhe-----AspValVal----- 113
Db 250 GCTGGAGACGTAGGACATCTCTTCAAAAAGCAATTTTCAGAATGGATGAGATGATGCAA 309
QY 114 ---GluArgSerPheLeuGluSerIleAspAlaLeuAlaGluLysAlaSerLeuGln 132
Db 310 GGACAAGAGGGTGGCGAGGTAGCA-----GTACTTGGTACAAAATCAATAAGTTC 363
QY 133 SerGlnLeuProGlu-GlyValProGlnHisGlnLeuProGlnTyrGlnLysIleLe 152
Db 364 AGTGGGATGATTGAAGGCTTATATGTGCACCA-----GAAGTGGGACAGTGTCT 414
QY 152 uGluArgLeu-----LysThrLeuGluArgGluIleSerGly----- 164
Db 415 AATAAACCTGATGCTTGGCGCTTTGAGGAAGGCTCTCAT-TCTGATTTTCTGTGACCTAA 473
QY 165 ---GlyAlaMetAlaValAlaValLeuLeuAsnLysLeuTyrValAlaAsnVal 183
Db 474 TTCTGGGACACGGCATCGTAGCTGTGTTAGAGACACACGACTATTTGTCGAATGC 533
QY 183 lGlyThrAsnArgAlaLeuLeuCysLysSerThrValAspGlyLeuGlnValThrGlnLe 203
Db 534 AGGTGACTCACGCTGTGTGATATCCAGAAAGAAAT-----CAGGCTTATATATCT 581
QY 203 uAsnValAspHisThrThrGluAsnGluAspGluLeuPheArgLeuSerGlnLysLe 223
Db 1147 TGCACACGCCCCACT 1161
```

```
Db 582 TTCTAGATCATCAAAACAGATCTTGAAGCTGAGAAAGAAAGATA-----TT 629
Qy 223 UASPAIAGLYLVYIILEYSGINVAIGLYILEYSGLYGIN---GlusterThargar 242
Db 630 GAAAGCTGGGGCTTATATCATGACGGGGGAGTCAATGAAGCTTAATCATATACGAGC 669
Qy 242 gIleGIAspTYrLYsValIySTyrcIyThraspIleAspLeuSerAlaVal 262
Db 690 TATCGGAGCATGGAATTCAGACAG-----ATAAGTTTGGCATCTGA 734
Qy 262 sSerLYsProIleIleAlaGluProGluIleHisGlyAlaGlnProleuAspGlyValTh 282
Db 735 AAGACCAATAGTTACCGCTAGTCCAGATGTATACCTGTGAAGCTGTGTGAT--GA 791
Qy 282 rGIyPheLeuValIleuMetSerGluIyLeuTYrLYsAlaLeuGluIleAlaHisGlyPr 302
Db 792 TGATTTCTGTTCTTGGCTGCGATGGAATTTGGGATGGCATG----- 834
Qy 302 oGIyGlnAlaAsnGlnGluIleAlaIleMetIleAspThrGluPheAlaLYsGlnThrSe 322
Db 835 -----ACAAGCCCAACACTCGTTGATTTCAATCATGAAACATTGAATTCAGAGACCA 887
Qy 322 rLeuAspAlaValAlaGlnAlaValAlaAspArgValLYsArgIleHisSerAspThrPh 342
Db 888 ACTCGCGGTGTATGTGAAGAAAGTTCTCGATGATGTGCGCTCCAAACACT----- 929
Qy 342 eAlaSerGIyGIyGluArgAlaArgPheCYsProArgHisGluAspMetThrLeuLeuVa 362
Db 940 -----TCAGGGGGTGAAGGCTGT-----GATAACAGAACCAATGATATT 977
Qy 362 lArgAsnPhGIyTYrProleuGIyGluMetSerGlnProThrProSerProAlaProAl 382
Db 978 GCTTCGATTCAAGAACCTCTCATCCAGACCAACCAACCAAGACCAAGACCGAGC 1037
Qy 382 aAlaGIyGIyArgValTYrProValSer 391
Db 1038 AGAAGGAAACCAAGATGAGCCGAGCTCA 1065

RESULT 15
US-09-770-445-6
; Sequence 6, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Xu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Wessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hubban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1496
```

```
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1496)
; OTHER INFORMATION: n = A,T,C or G
US-09-770-445-6

Alignment Scores:
Pred. No.: 2,61e-08 Length: 1496
Score: 166.00 Matches: 97
Percent Similarity: 42.39% Conservative: 70
Best Local Similarity: 24.62% Mismatches: 161
Query Match: 7.21% Indels: 67
DB: 10 Gaps: 17

US-09-830-144-4 (1-504) x US-09-770-445-6 (1-1496)
Qy 21 leuProleuCYs-HisLeuSerGIy-----ValGlySerAlaSe 33
Db 131 CTTCCTTTGTTTCATTTCTTACCCAGATATCATGGGTACATACCTAAGTTCTCCGAA 190
Qy 33 rAsnArgSerTYrSerAlaAspGIyLySGIyThrGluSerHisProProGluAspSerTr 53
Db 191 AACTGAAAGTTATCGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 250
Qy 53 pLeuLYsPheArgSer-----GluAsnAs 61
Db 251 GCAAAGTTGGGGCTTACCATGGAAGACCGCATGCTGCATTTCTTGATCTTGATGATAA 310
Qy 61 nCYsPheLeuTYrGIyValPheAsnGIyTYrAspGIyAsnArgValThrasnPhValAl 81
Db 311 GACATCGTTCGTCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 370
Qy 81 aGlnArgLeuSerAlaGluLeuLeuLeuGIyGln---LeuAsnAlaGlnHisAlaGlnAl 100
Db 371 CAAGATCTACACCGACGGGTATTCAGTATGAGCGTATTAATTCGAGACGTCGAAAC 430
Qy 100 aAspValArgArgValLeuLeuGlnAlaPheAspValAl-----GluArgSerPheLe 118
Db 431 ATCTCTTGAAGACATCTTTAGAAATGATGATGATGATGATGATGATGATGATGATGAT 490
Qy 118 uGluSerIleAspAspAlaGluLeuAlaGluLYsAlaSerLeuGlnSerGlnLeuProGlu 138
Db 491 AGAGTTAGCT-----GTACTTGGCAGCAAGATGAACAATTTAGCGGCATGATGAAGG 544
Qy 138 Y-ValProGlnHisGln-----LeuProProGlnIleGlnLYsIleuGluArgLeu 156
Db 545 ATTTATATGCTCACCAAGACGGGTGACCAATTAACCAACCCGATAGTGGCTCTTGA 604
Qy 156 YG-----ThLeuGluArgGluIleSerGIyLYsAlaMetAlaValAlaVal 172
Db 605 AGATGCTTCATTTCTGATTCACGGGACCTACCTCGGGGTGACCA-CCGTTGTAGGCTC 663
Qy 172 aLeuLeuAsnAsnLYsLeuTYrValAlaAsnValGIyThrasnArgAlaLeuLeuCYsL 192
Db 664 TTATTAAAGATTAAGACACTCTTTCTTGCAAAATCCGGTACCTAGTGTGATATCAAA 723
Qy 192 ySerThrValAspGIyLeuGlnValThrGlnLeuAsnValAspHisThrThrGluAsn 212
Db 724 GAAAGAT-----CAGGCTTACATCTTTCTAAATATCAAGACCTGATCTTG 771
Qy 212 lAspGluLeuPheArgLeuSerGlnLeuGIyLeuAspAlaGIyLYsIleLYsGlnValG 232
Db 772 AAGTTGAAAGAAAGAAAGATA-----TTGAACTGGTGGCTTTATTCACGCTG 819
Qy 232 lYlleIleCYsGIyGln---GluSerThArgArgIleGIyAspTYrLYsValLYsTYrG 251
Db 820 GGAGAAATCAATGAAGCTTGAATCTGACAAAGACCTTGGTATATGAGATTCAGACG- 878
Qy 251 lYTYrThraspIleAspLeuLeuSerAlaAlaLYsSerLYsProIleIleAlaGluPro 271
Db 879 -----ATAAGTTTTCACATCTGAAAGCAATGATTACTGCTATCCAG 924
```

```
Qy 271 luileHisGlyAlaGlnProLeuAspGlyValThrGlyPheLeuValLeuMetSerGluG 291
Db 925 ATATAAACACTATTGACCTATGTGATGAT--GATGACTTCTTGTGTTNNNNNNNTN 981
Qy 291 lYLeuTYrLYsAlaLeuGluAlaAlaHisGlyProGlyGlnAlaAsnGlnLileAlaA 311
Db 982 NNANATGGGATTGTATG-----TCAAGCCAGGAACTAGTTG 1017
Qy 311 laMetIleAspThrGluPheAlaLYsGlnThrSerLeuAspAlaValAlaGlnAlaValV 331
Db 1018 ATTTTATCCATGAACAGTTAAATCTGAACAAACACTTTCAACAGTATGTCAAAAGGTTG 1077
Qy 331 alaSpargValLYsArgIleHisSerAspThrPheAlaSerGlyGlyGluArgAlaArgP 351
Db 1078 TTGATAGATGT-----TTGGCTCCAGATACAGCCGACTGGTGAAGGT----- 1118
Qy 351 heCysProArgHisGluAspMetThrLeuLeuValArgAsnPheGlyTYrProLeuGlyG 371
Db 1119 --TGT-----GATAATATGACCATCATCTTGTTCAGTTCAAGAAGCCTAACCCAT 1167
Qy 371 luMetSerGlnProThrProSer---ProAlaProAla 382
Db 1168 CTGAGACTGAACCAAGAAGATTCCAACCAAGACCAAGC 1205
```

Search completed: December 10, 2002, 02:09:31  
Job time : 98.5983 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 05:57:28 : Search time 85 Seconds  
(without alignments)  
106.600 Million cell updates/sec

Title: US-09-830-144-4\_COPY\_437\_504

Perfect score: 359  
Sequence: 1 OSPRLTQSTWTHQSSSSS.....AEFRYSVDHGEQSVTAP 68

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq.101002:\*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	359	100.0	504	18	AAW26706
2	359	100.0	504	18	AAW26707
3	359	100.0	504	20	AAW09541
4	359	100.0	504	21	AAV91001
5	359	100.0	504	21	AAV59450
6	359	100.0	513	20	AAV09550
7	359	100.0	517	20	AAV09546
8	352	98.1	84	21	AAW56692
9	77	21.4	16	20	AAV09549
10	70	19.5	22	22	ABB70838

11	69.5	19.4	199	22	AAV78804	Human HT4SG64, ser1
12	68.5	19.1	113	22	AAV47075	Protonibacterium
13	67.5	18.8	739	22	ABG16477	Novel human diageno
14	67.5	18.8	739	22	AAV68173	Novel human diageno
15	67.5	18.8	852	20	AAV30948	Atrophin-1, interac
16	66	18.4	16	20	AAV09548	Human E3 ubiquitin
17	64.5	18.0	116	20	AAV09548	Human TAB1 peptide
18	64.5	18.0	244	21	AAV65525	Novel human secret
19	63.5	17.7	187	21	AAV56555	Human gene 72-enco
20	63.5	17.7	219	21	AAV35654	Arabidopsis thalia
21	63.5	17.7	240	19	AAV75243	Arabidopsis thalia
22	63.5	17.7	241	22	ABG03626	Fragment of human
23	63.5	17.7	243	21	AAV35653	Novel human diageno
24	63.5	17.7	335	19	AAV75208	Arabidopsis thalia
25	63.5	17.7	391	22	AAV36176	Human secreted pro
26	63.5	17.7	824	22	AAV20251	Human DIB2 protein
27	63.5	17.7	824	23	AAV86156	Human ADAM6, Homo
28	63	17.5	66	22	AAV65430	Human PDI686 poly
29	63	17.5	2112	22	AAV60403	Human immune/thema
30	62.5	17.4	92	21	AAV42825	Drosophila melanog
31	62.5	17.4	675	22	AAV50204	Human ORFX ORF2589
32	62.5	17.4	1945	22	ABG64947	Protonibacterium
33	62	17.3	527	19	AAV48419	Drosophila melanog
34	62	17.3	734	22	AAV82316	Amino acid sequenc
35	62	17.3	3583	22	ABG64814	Human immunoglobul
36	61	17.0	60	22	AAV65449	Drosophila melanog
37	61	17.0	322	22	ABG10298	Protonibacterium
38	61	17.0	1511	21	AAV28182	Human cDNA SEQ ID
39	60.5	16.9	854	20	AAV30949	Yeast transporter
40	60	16.7	1037	20	AAV67643	Murine E3 ubiquiti
41	59	16.4	173	23	ABG55244	A serine/threonine
42	59	16.4	1046	22	AAV11775	Lactococcus lactis
43	59	16.4	1327	22	ABG58539	Human kinase (PKIN
44	58.5	16.3	90	22	AAV56766	Drosophila melanog
45	58.5	16.3	1093	16	AAV66460	Protonibacterium

## ALIGNMENTS

RESULT 1	
ID	AAW26706 standard; Protein: 504 AA.
XX	
AC	AAW26706;
XX	
DT	14-APR-1998 (first entry)
XX	
DE	Human TAB1 (TAK1 binding protein).
XX	
KW	TAB1; TAK1 binding protein; transforming growth factor-beta;
XX	signal transduction; human.
OS	Homo sapiens.
XX	
FT	Key
FT	Misc-difference 52
XX	/note= "variant has Arg as residue 52"
XX	
PN	EB803571-AA.
XX	
PD	29-OCT-1997.
XX	
PF	24-APR-1997; 97EP-0302808.
XX	
PR	20-NOV-1996; 96US-0752891.
PR	24-APR-1996; 96JP-0136282.
PR	28-OCT-1996; 96JP-0300856.
XX	
PA	(UENO/) UENO N.
XX	
PI	Matsumoto K, Nishida E;
XX	

DR WPI; 1997-515318/48.  
DR N-PSDB; AAT91175.  
XX  
PT DNA encoding TAK1 binding protein TAB1 - member of transforming  
PT growth factor beta receptor signal production pathway, which  
PT activates TAK-1 kinase activity upon binding  
XX  
PS Example 5; Page 17-19; 30pp; English.  
XX  
CC This protein comprises human TAB1, a novel member of the  
CC transforming growth factor-beta receptor signal transduction  
CC pathway, which activates TAK-1 kinase activity upon binding. Its  
CC amino acid sequence was deduced from a cDNA clone (see AAT91175)  
CC obtained from a kidney library; a variant TAB1 (see AAW26707) has  
CC Arg rather than Ser at amino acid position 52. Also claimed are:  
CC isolated DNA encoding a protein modified by a substitution,  
CC deletion and/or addition of 1 or more amino acids of the 504-residue  
CC TAB1 sequence; (2) DNA which can hybridise with the 1560 bp TAB1  
CC nucleic acid sequence; (3) isolated DNA encoding a protein  
CC comprising amino acids 21-579 or 437-504 of the 504 TAB1 sequence;  
CC (4) DNA encoding a fusion protein comprising an above protein or  
CC polypeptide; (5) expression vector comprising an above DNA; and  
CC (6) host cell, preferably a mammalian or yeast cell, transformed by  
CC the expression vector. Cells expressing TAB1 and TAK1 can be used  
CC to screen for TGF-beta signalling pathway inhibitors by contacting  
CC the cells with a test compound, and measuring the TAK1 kinase  
CC activity.  
XX  
SQ Sequence 504 AA;  
Query Match 100.0%; Score 359; DB 18; Length 504;  
Best Local Similarity 100.0%; Pred. No. 3.1e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QSPFLTLQSTNTHTQSSSSSDGGLFRSRPAHSLPPGEGRVEPYVDFAEFYRLWSVDHG 60  
DB 437 QSPFLTLQSTNTHTQSSSSSDGGLFRSRPAHSLPPGEGRVEPYVDFAEFYRLWSVDHG 496  
QY 61 EQSVVTAP 68  
DB 497 EQSVVTAP 504  
RESULT 2  
AAW26707  
XX AAW26707 standard; Protein; 504 AA.  
XX  
XX AAW26707;  
XX  
XX 14-APR-1998 (first entry)  
XX  
XX Human TAB1 (TAK1 binding protein).  
XX  
XX TAB1; TAK1 binding protein; transforming growth factor-beta;  
XX signal transduction; human.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 52  
FT /note= "variant has Ser as residue 52"  
XX  
XX EP803571-A2.  
XX  
XX 29-OCT-1997.  
XX  
XX 24-APR-1997; 97EP-0302808.  
XX  
XX 20-NOV-1996; 96US-0752891.  
XX 24-APR-1996; 96JP-0126282.  
XX 28-OCT-1996; 96JP-0300856.  
XX  
XX (UENO/) UENO N.  
XX PA

XX  
PI Matsumoto K, Nishida E;  
XX  
DR WPI; 1997-515318/48.  
DR N-PSDB; AAT91178.  
XX  
PT DNA encoding TAK1 binding protein TAB1 - member of transforming  
PT growth factor beta receptor signal production pathway, which  
PT activates TAK-1 kinase activity upon binding  
XX  
PS Example 5; Page 19-21; 30pp; English.  
XX  
CC This protein comprises human TAB1, a novel member of the  
CC transforming growth factor-beta receptor signal transduction  
CC pathway, which activates TAK-1 kinase activity upon binding. Its  
CC amino acid sequence was deduced from a cDNA clone (see AAT91178)  
CC obtained from a kidney library; a variant TAB1 (see AAW26706) has  
CC Ser rather than Arg at amino acid position 52. Also claimed are:  
CC isolated DNA encoding a protein modified by a substitution,  
CC deletion and/or addition of 1 or more amino acids of the 504-residue  
CC TAB1 sequence; (2) DNA which can hybridise with the 1560 bp TAB1  
CC nucleic acid sequence; (3) isolated DNA encoding a protein  
CC comprising amino acids 21-579 or 437-504 of the 504 TAB1 sequence;  
CC (4) DNA encoding a fusion protein comprising an above protein or  
CC polypeptide; (5) expression vector comprising an above DNA; and  
CC (6) host cell, preferably a mammalian or yeast cell, transformed by  
CC the expression vector. Cells expressing TAB1 and TAK1 can be used  
CC to screen for TGF-beta signalling pathway inhibitors by contacting  
CC the cells with a test compound, and measuring the TAK1 kinase  
CC activity.  
XX  
SQ Sequence 504 AA;  
Query Match 100.0%; Score 359; DB 18; Length 504;  
Best Local Similarity 100.0%; Pred. No. 3.1e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QSPFLTLQSTNTHTQSSSSSDGGLFRSRPAHSLPPGEGRVEPYVDFAEFYRLWSVDHG 60  
DB 437 QSPFLTLQSTNTHTQSSSSSDGGLFRSRPAHSLPPGEGRVEPYVDFAEFYRLWSVDHG 496  
QY 61 EQSVVTAP 68  
DB 497 EQSVVTAP 504  
RESULT 3  
AAW09541  
XX AAW09541 standard; Protein; 504 AA.  
XX  
XX AAW09541;  
XX  
XX 21-JUL-1999 (first entry)  
XX  
XX Human TAB1 protein.  
XX  
XX Human; TAB1; TAK1; screening; inhibition; TGF-beta;  
XX transforming growth factor beta.  
XX  
XX Homo sapiens.  
XX  
XX WO9921010-A1.  
XX  
XX 29-APR-1999.  
XX  
XX 22-OCT-1998; 98WO-JP04796.  
XX  
XX 22-OCT-1997; 97JP-0290189.  
XX  
XX (CHUS ) CHUGAI SEIYAKU KK.  
XX  
XX Ohtomo T, Ono K, Tsuchiya M;  
XX

DR WPI; 1999-312645/26.  
DR N-PSDB; AAX56278.  
XX Screening for TGF- beta inhibitory substances, which are useful as  
PT drugs for treatment of diseases relating to its disorder  
XX  
XX Claim 3; Page 147-149; 195pp; Japanese.  
XX  
XX A method has been developed for screening for substances which inhibit  
CC the binding of TAK1 polypeptide to TAB1 polypeptide. The method  
CC comprises: (a) contacting the polypeptide in the presence of a sample;  
CC and (b) detecting the amount of bound polypeptide, in which the sample  
CC can be pre-mixed with TAK1 or TAB1 polypeptide first. The transforming  
CC growth factor (TGF)-beta inhibitory substances can be used in drugs for  
CC indications e.g. as TGF-beta signal transduction inhibitors or  
CC activators, or extracellular matrix protein production enhancement  
CC inhibitors, or activators, or cell proliferation prevention inhibitors or  
CC activators, or monocyte migration inhibitors or activators, or  
CC physiological activity induction inhibitors or activators, or  
CC immunosuppression inhibitors or activators, or amyloid beta protein  
CC precipitation inhibitors or activators, and such substances can also be  
CC inhibitors of the TAK1 polypeptide function, particularly kinase  
CC activity. The present sequence represents human TAB1.  
CC  
XX  
SQ Sequence 504 AA;  
Query Match 100.0%; Score 359; DB 20; Length 504;  
Best Local Similarity 100.0%; Pred. No. 3.1e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QSPFLTLQSTNTHQSSSSSDGGLFRSRPAHSLEPPGDEGRVEPYVDFAFYRLMSVDHG 60  
DB 437 QSPFLTLQSTNTHQSSSSSDGGLFRSRPAHSLEPPGDEGRVEPYVDFAFYRLMSVDHG 496  
QY 61 EQSVVTAP 68  
DB 497 EQSVVTAP 504  
Db  
RESULT 4  
AAV91001  
ID AAY91001 standard; protein; 504 AA.  
XX  
AC AAY91001;  
XX  
DT 04-SEP-2000 (first entry)  
XX  
DE Human TAB-1 protein sequence SEQ ID NO:4.  
XX  
XX Human; TAK-1; TAB-1; mitogen activated protein kinase; MAPK;  
KW screening; signal transduction; inhibition; inflammatory cytokine;  
KW IL-1; interleukin 1; TNF; tumour necrosis factor; inflammation;  
KW antiinflammatory; suppression.  
XX  
OS Homo sapiens.  
XX  
PN WO200023610-A1.  
XX  
PD 27-APR-2000.  
XX  
PF 21-OCT-1999; 99WO-JP05817.  
XX  
PR 21-OCT-1998; 98JP-0299962.  
XX  
PA (CHUS ) CHUGAI SEIYAKU KK.  
XX  
PI Tsuchiya M, Ohtomo T, Sugamata Y, Matsumoto K;  
XX  
DR WPI; 2000-339707/29.  
DR N-PSDB; AAA39106.  
PT Method for screening inhibitors of TAK1 signal transduction for  
suppression of inflammatory cytokine production and use as

PT antiinflammatory agents -  
XX  
XX Disclosure; Page 90-94; 100pp; Japanese.  
XX  
XX The present invention describes a method for screening compounds for  
CC inhibition of inflammatory cytokine signal transduction by contacting  
CC the sample with TAK1 and its receptor TAB1 and selecting for inhibition  
CC of TAK1/TAB1 binding. Also described is a method for screening compounds  
CC for inhibition of inflammatory cytokine signal transduction in which the  
CC inhibition of TAK1 phosphorylation is selected for; and drug  
CC compositions for the treatment of inflammatory disorders containing as  
CC active component an inflammatory cytokine signal transduction inhibitor.  
CC TAK1 is an essential component of the signalling process which results  
CC in release of inflammatory cytokines such as interleukin-1 (IL-1).  
CC IL-10, tumour necrosis factor (TNF) and IL-6. The methods can be used  
CC for the selection of effective antiinflammatory agents. The present  
CC sequence represents human TAB-1, which is used in the exemplification of  
CC the present invention.  
CC  
XX  
SQ Sequence 504 AA;  
Query Match 100.0%; Score 359; DB 21; Length 504;  
Best Local Similarity 100.0%; Pred. No. 3.1e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QSPFLTLQSTNTHQSSSSSDGGLFRSRPAHSLEPPGDEGRVEPYVDFAFYRLMSVDHG 60  
DB 437 QSPFLTLQSTNTHQSSSSSDGGLFRSRPAHSLEPPGDEGRVEPYVDFAFYRLMSVDHG 496  
QY 61 EQSVVTAP 68  
DB 497 EQSVVTAP 504  
Db  
RESULT 5  
AAV59450  
ID AAV59450 standard; Protein; 504 AA.  
XX  
AC AAV59450;  
XX  
DT 24-MAR-2000 (first entry)  
XX  
DE Human TAB1 protein sequence.  
XX  
XX Human; TAB1; XIAP; X-linked inhibitor of apoptosis protein; TGF-beta;  
KW transforming growth factor-beta activated kinase 1; monocyte migration;  
KW TAK1 binding protein 1; extracellular matrix protein production;  
KW cell growth inhibitor; beta-amyloid protein deposition;  
KW immunosuppression; Transforming growth factor-beta.  
XX  
OS Homo sapiens.  
XX  
PN JP1326328-A.  
XX  
PD 26-NOV-1999.  
XX  
PF 13-MAY-1998; 98JP-0130378.  
XX  
PR 13-MAY-1998; 98JP-0130378.  
XX  
PA (MATS/) MATSUMOTO K.  
XX  
DR WPI; 2000-078337/07.  
DR N-PSDB; AA248861.  
PT Screening a substance which inhibits combination of the X-linked  
XX inhibitor of apoptosis protein -  
XX  
XX Claim 2; Page 25-26; 43pp; Japanese.  
XX  
XX This sequence represents the human TAB1 protein.  
CC The invention relates to a method for screening a substance inhibiting  
CC the formation of a complex between XIAP and TAB1, in which X-linked

CC inhibitor of apoptosis protein (XIAP), transforming growth factor-beta  
 CC activated kinase 1(TAK1) binding protein 1(TAB1) and a substance to be  
 CC tested are contacted with each other and then the presence or formation  
 CC of a complex between XIAP and TAB1 is detected. The substance can be used  
 CC as a drug for extracellular matrix protein production enhancement, cell  
 CC growth inhibition, monocyte migration, physiologically active substance  
 CC induction, immunosuppression, and beta-amyloid protein deposition. A  
 CC substance inhibiting the formation of a complex between TAB1 and XIAP as  
 CC well as between XIAP and TGF-beta (transforming growth factor-beta) type  
 CC I and/or type II receptor is useful as a drug.

Query Match 100.0%; Score 359; DB 21; Length 504;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-37;  
 Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QSTPLTQSTNTHTQSSSSDGLFRSPAHSLPPGDRGVRPEYVDFAEFRLWSVDHG 60  
 DB 437 QSTPLTQSTNTHTQSSSSDGLFRSPAHSLPPGDRGVRPEYVDFAEFRLWSVDHG 496

QY 61 EQSVVTAP 68  
 DB 497 EQSVVTAP 504

RESULT 6  
 AAY09550  
 ID AAY09550 standard; Protein; 513 AA.

XX AC AAY09550;  
 XX 21-JUL-1999 (first entry)  
 XX Human TAB1 protein SEQ ID NO:43.  
 XX Human; TAB1; TAK1; screening; inhibition; TGF-beta;  
 XX transforming growth factor beta.

XX OS Homo sapiens.  
 XX WO9921010-A1.  
 XX 29-APR-1999.  
 XX 22-OCT-1998; 98WO-JP04796.  
 XX 22-OCT-1997; 97JP-0290188.  
 XX (CHUS ) CHUGAI SEIYAKU KK.

XX Ontomo T, Ono K, Tsuchiya M;  
 XX WPI; 1999-312645/26.  
 XX N-PSDB; AAX56310.

PT Screening for TGF- beta inhibitory substances, which are useful as  
 PT drugs for treatment of diseases relating to its disorder

PS Example 13; Page 186-188; 195pp; Japanese.

XX A method has been developed for screening for substances which inhibit  
 CC the binding of TAK1 polypeptide to TAB1 polypeptide. The method  
 CC comprises: (a) contacting the polypeptide in the presence of a sample;  
 CC and (b) detecting the amount of bound polypeptide, in which the sample  
 CC can be pre-mixed with TAK1 or TAB1 polypeptide first. The transforming  
 CC growth factor (TGF)-beta inhibitory substances can be used in drugs for  
 CC indications e.g. as TGF-beta signal transmission inhibitors or  
 CC activators, or extracellular matrix protein production enhancement  
 CC inhibitors, or activators, or cell proliferation prevention inhibitors or  
 CC activators, or monocyte migration inhibitors or activators, or  
 CC physiological activity induction inhibitors or activators, or  
 CC immunosuppression inhibitors or activators, or amyloid beta protein

CC precipitation inhibitors or activators, and such substances can also be  
 CC inhibitors of the TAK1 polypeptide function, particularly kinase  
 CC activity. The present sequence represents human TAB1.

QY Sequence 513 AA;

Query Match 100.0%; Score 359; DB 20; Length 513;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-37;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QSTPLTQSTNTHTQSSSSDGLFRSPAHSLPPGDRGVRPEYVDFAEFRLWSVDHG 60  
 DB 446 QSTPLTQSTNTHTQSSSSDGLFRSPAHSLPPGDRGVRPEYVDFAEFRLWSVDHG 505

QY 61 EQSVVTAP 68  
 DB 506 EQSVVTAP 513

RESULT 7  
 AAY09546  
 ID AAY09546 standard; Protein; 517 AA.

XX AC AAY09546;  
 XX 21-JUL-1999 (first entry)  
 XX Human TAB1-FLAG protein.  
 XX Human; TAB1; TAK1; screening; inhibition; TGF-beta;  
 XX transforming growth factor beta.

XX OS Homo sapiens.  
 XX Synthetic.  
 XX WO9921010-A1.  
 XX 29-APR-1999.

XX 22-OCT-1998; 98WO-JP04796.  
 XX 22-OCT-1997; 97JP-0290188.  
 XX (CHUS ) CHUGAI SEIYAKU KK.

XX Ontomo T, Ono K, Tsuchiya M;  
 XX WPI; 1999-312645/26.  
 XX N-PSDB; AAX56282.

PT Screening for TGF- beta inhibitory substances, which are useful as  
 PT drugs for treatment of diseases relating to its disorder

PS Example 1; Page 163-166; 195pp; Japanese.

XX A method has been developed for screening for substances which inhibit  
 CC the binding of TAK1 polypeptide to TAB1 polypeptide. The method  
 CC comprises: (a) contacting the polypeptide in the presence of a sample;  
 CC and (b) detecting the amount of bound polypeptide, in which the sample  
 CC can be pre-mixed with TAK1 or TAB1 polypeptide first. The transforming  
 CC growth factor (TGF)-beta inhibitory substances can be used in drugs for  
 CC indications e.g. as TGF-beta signal transmission inhibitors or  
 CC activators, or extracellular matrix protein production enhancement  
 CC inhibitors, or activators, or cell proliferation prevention inhibitors or  
 CC activators, or monocyte migration inhibitors or activators, or  
 CC physiological activity induction inhibitors or activators, or  
 CC immunosuppression inhibitors or activators, or amyloid beta protein  
 CC precipitation inhibitors or activators, and such substances can also be  
 CC inhibitors of the TAK1 polypeptide function, particularly kinase  
 CC activity. The present sequence represents TAB1-FLAG from an example of  
 CC the present invention.

XX Sequence 517 AA;

Query Match 100.0%; Score 359; DB 20; Length 517;  
Best Local Similarity 100.0%; Pred. No. 3.2e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 OSPTLTLOSTNTHTQSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDPAEFYRLMSVDHG 60  
Db 437 OSPTLTLOSTNTHTQSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDPAEFYRLMSVDHG 496  
Qy 61 EOSVVTAP 68  
Db 497 EOSVVTAP 504

## RESULT 8

AAB56692 ID AAB56692 standard; Protein; 84 AA.

AC AAB56692;

DT 13-MAR-2001 (first entry)

DE Human prostate cancer antigen protein sequence SEQ ID NO:1270.

Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
neuroprotective; cytoskeletal; cardiolipin; immunomodulatory; muscular;  
vulnerable; gastrointestinal; nephrotoxic; antileukemic; gynaecological;  
antibacterial; gene therapy; neutral; immune; reproductive; renal;  
gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
wound; infectious disease.

OS Homo sapiens.

PN WO20005174-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05988.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

PI Rosen CA, Ruben SM;

DR WPI; 2000-587513/55.

DR N-PSDB; AAF15895.

PT Prostate cancer associated gene sequences, referred to as prostate  
cancer antigens, useful for treatment, prevention, and diagnosis of  
disorders such as prostate cancer -

PS Claim 11; Page 1693; 2338pp; English.

AAPI5566 to AAPI5505 encode the human prostate cancer associated  
proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
The prostate cancer antigens can have neuroprotective, cytoskeletal,  
cardioactive, immunomodulatory, muscular, vulnerable, gastrointestinal,  
neurotropic, antileukemic, gynaecological and antibacterial activities,  
and can be used in gene therapy. The prostate cancer antigen  
polynucleotides may be used for detection of prostate cancer, chromosome  
identification, as chromosome markers, and for numerous other diagnostic  
or research purposes. The prostate cancer antigens may be used to treat  
disorders such as neural, immune, muscular, reproductive,  
gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
disorders, wounds, and infectious diseases. AAPI5506 to AAPI5514 to  
CC AAB57303 represent sequences used in the exemplification of the present  
invention.

SQ Sequence 84 AA;

Query Match

98.1%; Score 352; DB 21; Length 84;

Best Local Similarity 98.5%; Pred. No. 2.4e-37;  
Matches 67; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 OSPTLTLOSTNTHTQSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDPAEFYRLMSVDHG 60  
Db 17 OSPTLTLOSTNTHTQSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDPAEFYRLMSVDHG 76  
Qy 61 EOSVVTAP 68  
Db 77 EOSVVTAP 84

## RESULT 9

AAAY09549 ID AAAY09549 standard; peptide; 16 AA.

AC AAAY09549;

DT 21-JUL-1999 (first entry)

DE Human TAB1 peptide TAB1C-2.

Human; TAB1; screening; inhibition; TGF-beta;  
transforming growth factor beta.

OS Homo sapiens.

PN WO9921010-A1.

PD 29-APR-1999.

PF 22-OCT-1998; 98WO-JP04796.

PR 22-OCT-1997; 97JP-0290188.

PA (CHUS) CHUGAI SEIYAKU KK.

PI Ohtomo T, Ono K, Tsuchiya M;

DR WPI; 1999-312645/26.

PT Screening for TGF-beta inhibitory substances, which are useful as  
drugs for treatment of diseases relating to its disorder

PS Example 13; Page 182; 195pp; Japanese.

A method has been developed for screening for substances which inhibit  
the binding of TAK1 polypeptide to TAB1 polypeptide. The method  
comprises: (a) contacting the polypeptide in the presence of a sample;  
and (b) detecting the amount of bound polypeptide, in which the sample  
can be pre-mixed with TAK1 or TAB1 polypeptide first. The transforming  
growth factor (TGF)-beta inhibitory substances can be used in drugs for  
indications e.g. as TGF-beta signal transduction inhibitors or  
activators, or extracellular matrix protein production enhancement  
inhibitors or activators, or cell proliferation prevention inhibitors or  
activators, or monocyte migration inhibitors or activators, or  
physiological activity induction inhibitors or activators, or  
immunoppression inhibitors or activators, or amyloid beta protein  
precipitation inhibitors or activators, and such substances can also be  
inhibitors of the TAK1 polypeptide function, particularly kinase  
activity. The present sequence represents a peptide from an example  
of the present invention.

SQ Sequence 16 AA;

Query Match 21.4%; Score 77; DB 20; Length 16;

Best Local Similarity 100.0%; Pred. No. 0.0046;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 OSPTLTLOSTNTHTQ 15  
Db 2 OSPTLTLOSTNTHTQ 16

```
RESULT 10
ABB70838
ID ABB70838 standard; Protein; 1162 AA.
XX
XX ABB70838;
AC
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 39306.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW
KW Drosophila melanogaster.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
PR
XX (PEKE ) PE CORP NY.
XX
PA Venter JC, Adams M, Li PWD, Myers BW;
XX
PI WPI; 2001-656860/75.
XX
DR N-PSDB; ABL14941.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
XX Disclosure; SEQ ID NO 39306; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLL16176-ABLL30511), expressed DNA
XX sequences (ABLL01840-ABLL16175) and the encoded proteins
XX (ABBS7737-ABBS72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1162 AA;

Query Match 19.5%; Score 70; DB 22; Length 1162;
Best Local Similarity 32.8%; Pred. No. 9.8;
Matches 22; Conservative 9; Mismatches 24; Indels 12; Gaps 2;

QY 3 PTLTLOSTWTHQSSSSSDGGLFRS-RPAHSI.PGCDGRVPEYVDFAEFYRLMSVDHGE 61
Db 960 PVLQPKSPSTTLSCSSSSAGSAGAYQYAPGMLPPR-----ADVRLSLSLNGS 1008
QY 62 QSVVTAP 68
Db 1009 SSEVTSP 1015

RESULT 11
AAG77804
ID AAG77804 standard; Protein; 199 AA.
XX
XX AAG77804;
AC
XX 03-DEC-2001 (first entry)
DT
XX
```

DE Human HT4SG64 serine/threonine phosphatase protein sequence.

XX Human; HT4SG64; serine/threonine phosphatase; PSPase; vaccine; gene therapy; PSPase expression; PSPase expression; PSPase modulation; immune disorder; autoimmune disorder; Wiscott-Aldrich syndrome; Chediak-Higashi syndrome; Hashimoto's thyroiditis; multiple sclerosis; inflammation; Crohn's disease; inflammatory bowel disease; appendicitis; rheumatoid arthritis; cellular proliferative disorder; lymphoma; lung cancer; intestinal cancer; cardiovascular disorder; aneurysm; Scimitar syndrome; Ebstein's anomaly.

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT Region 70..84

FT /note= "Immunogenic region 1"

FT Region 114..121

FT /note= "Immunogenic region 2"

FT Region 126..136

FT /note= "Immunogenic region 3"

FT Region 159..164

FT /note= "Immunogenic region 4"

FT Region 186..199

FT /note= "Immunogenic region 5"

XX

PN WO200164703-A1.

XX

XX 07-SEP-2001.

PD

XX 28-FEB-2001; 2001WO-US06256.

PF

XX 02-MAR-2000; 2000US-0186350.

PR

XX (HUMA-) HUMAN GENOME SCI INC.

XX

PA Ebner R, Ruben SM;

XX

PI WPI; 2001-530113/58.

XX

DR N-PSDB; AAH78731.

DR

XX Nucleic acids encoding serine/threonine phosphatase polypeptides, useful for preventing, diagnosing and/or treating, e.g. Crohn's disease, lung cancer and Scimitar syndrome -

PT

PT

PT

XX

PS Claim 11; Page 323-324; 335pp; English.

XX

CC The present sequence represents the human HT4SG64 protein which is claimed in the invention. The invention comprises novel human serine/threonine phosphatase (PSPase) polypeptides and polynucleotides. The PSPase polynucleotides and polypeptides of the invention may be used in the prevention (vaccine), diagnosis and treatment (gene therapy) of diseases associated with inappropriate PSPase expression. The PSPase polynucleotides of the invention may be used as DNA probes to detect and quantitate the presence of similar nucleic acids in samples. The PSPase polypeptides may be used as antigens in the production of antibodies against the PSPase polypeptides and in assays to identify modulators of PSPase expression and activity. The anti-PSPase antibodies and antagonists may also be used to down regulate expression and activity, the anti-PSPase antibodies may also be used as diagnostic agents for detecting the presence of PSPase polypeptides in samples. Disorders that may be prevented, diagnosed and/or treated by the invention are: immune/autoimmune disorders (e.g. Wiscott-Aldrich syndrome, Chediak-Higashi syndrome, Hashimoto's thyroiditis and multiple sclerosis); inflammatory conditions (e.g. Crohn's disease, inflammatory bowel disease, appendicitis and rheumatoid arthritis); cellular proliferative disorders (e.g. lymphoma, lung and intestinal cancers); and cardiovascular disorders (e.g. Scimitar syndrome, Ebstein's anomaly and aneurysm).

XX

SQ Sequence 199 AA;

Query Match 19.4%; Score 69.5; DB 22; Length 199;

Best Local Similarity 41.9%; Pred. No. 1.1;

Matches 18; Conservative 3; Mismatches 17; Indels 5; Gaps 1;  
Qy 1 QSPFLTIGSTNTHTOSSSSSDGGLFRSRPAHSIPGDEGRVEYVDFAE 38  
| | | | | : | | | | | : | | | | | : | | | | | :  
| | | | | : | | | | | : | | | | | : | | | | | :  
Db 119 QSPFLTIGSTNTHTOSSSSSDGGLFRSRPAHSIPGDEGRVEYVDFAE 161

RESULT 12  
AAU47075  
ID AAU47075 standard; Protein; 113 AA.

AC AAU47075;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #7971.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KM dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhalaria A;

PI L'maisonmeuve J, Zhang Y, Jen S, Carter D;

DR WPI; 2001-616774/71.

DR N-PSDB; AASS9536.

PS Example 1; SEQ ID No 8270; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 113 AA;

Query Match 19.1%; Score 68.5; DB 22; Length 113;  
Best Local Similarity 35.4%; Pred. No. 0.72;

Matches 17; Conservative 8; Mismatches 22; Indels 1; Gaps 1;  
Qy 3 PTLTIGSTNTHTOSSSSSDGGLFRSRPAHSIPGDEGRVEYVDFAE 50  
| | | | | : | | | | | : | | | | | : | | | | | :  
| | | | | : | | | | | : | | | | | : | | | | | :  
Db 61 PTLTIGSTNTHTOSSSSSDGGLFRSRPAHSIPGDEGRVEYVDFAE 107

RESULT 13  
ABG16477  
ID ABG16477 standard; Protein; 739 AA.

AC ABG16477;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #16468.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.

OS WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS80664.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

PS Claim 20; SEQ ID No 46836; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 739 AA;

Query Match 18.8%; Score 67.5; DB 22; Length 739;  
Best Local Similarity 31.1%; Pred. No. 11;  
Matches 23; Conservative 12; Mismatches 20; Indels 19; Gaps 3;

```
QY 2 SPTLTLOSTNTHTQSSSSSDG-----GLFRSRPAH-----SLPPGDGRVEPYVDF 48
Db 119 SSTGSLPPTNTNTNSEGATSLIIPLTISGSGRFLNPVTQAPLPFGWEQRVD----- 173
QY 49 AEFYRLWSVDHGEQ 62
Db 174 -QHGRVYVDHVEK 186

RESULT 14
AAG68173
ID AAG68173 standard; Protein; 739 AA.
XX
AC AAG68173;
XX
DT 25-JAN-2002 (first entry)
XX
DE Atrophin-1 interacting protein (AIP4) SEQ ID NO:89.
XX
KW Human; high bone mass; HBM gene; Zmax1 gene; chromosome 11; 11q13.3;
KW sequence tagged site; STS; osteoporosis; osteopathic; gene therapy;
KW antisense therapy; vaccine; bone disorder; Paget's disease;
KW sclerostosis; Osteomalacia; fibrous dysplasia.
XX
OS Homo sapiens.
XX
PN WO200177327-A1.
XX
PD 18-OCT-2001.
XX
PF 21-JUN-2000; 2000WO-US16951.
XX
PR 05-APR-2000; 2000US-0543771.
XX
PR 05-APR-2000; 2000US-0544398.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Carulli JP, Little RD, Recker RR, Johnson ML;
XX
DR WPI; 2001-657171/75.
XX
XX New high bone mass (HBM) and Zmax1 genes and proteins useful for
PT modulating bone mass for the treatment of e.g. osteoporosis -
XX
PS Claim 76; Page 392-394; 443pp; English.
XX
CC The present invention describes the human Zmax1 gene and the high bone
CC mass (HBM) gene, which are found on chromosome 11q13.3. The Zmax1 and
CC HBM genes have osteopathic activities. The genes can be used in gene
CC therapy, antisense therapy and in the production of vaccines. They
CC can be used in the diagnosis and treatment of bone disorders including
CC osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous
CC dysplasia. ABA82038 to ABA82700 and AAG68168 to AAG68193 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 739 AA;

Query Match 18.8%; Score 67.5; DB 22; Length 739;
Best Local Similarity 31.1%; Pred. No. 11;
Matches 23; Conservative 12; Mismatches 20; Indels 19; Gaps 3;

QY 2 SPTLTLOSTNTHTQSSSSSDG-----GLFRSRPAH-----SLPPGDGRVEPYVDF 48
Db 119 SSTGSLPPTNTNTNSEGATSLIIPLTISGSGRFLNPVTQAPLPFGWEQRVD----- 173
QY 49 AEFYRLWSVDHGEQ 62
Db 174 -QHGRVYVDHVEK 186

RESULT 15
AAY30948
ID AAY30948 standard; Protein; 852 AA.
```

```
XX
AC AAY30948;
XX
DT 21-OCT-1999 (first entry)
XX
DE Human E3 ubiquitin protein ligase protein.
XX
KW E3 ubiquitin protein ligase; h-E3 UPL; antiinflammatory; antidiabetic;
KW immunosuppressive; neuroprotective; cytosolic; antiarthritic; cardiac;
KW immunomodulator; antiviral; treatment; screening; gene therapy; cancer;
KW inflammation; autoimmune disease; neurological disease; apoptosis;
KW endothelial cell; proliferation; differentiation; angiogenesis; cachexia;
KW peripheral vascular disease; hematopoietic disorder; arthritis; leukemia;
KW pulmonary disorder; diabetes; viral infection; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 275..306
FT /label= WWI
FT /note= "WW protein interaction domain I"
FT Domain 307..340
FT /label= WWII
FT /note= "WW protein interaction domain II"
FT Domain 386..420
FT /label= WWIII
FT /note= "WW protein interaction domain III"
FT Domain 427..460
FT /label= WWIV
FT /note= "WW protein interaction domain IV"
XX
PN WO9940201-A1.
XX
PD 12-AUG-1999.
XX
PF 02-FEB-1999; 99WO-GB00353.
XX
PR 30-APR-1998; 98US-0070060.
XX
PR 05-FEB-1998; 98US-0073839.
XX
PA (ZENE ) ZENECA LTD.
XX
PI Childyal N, Hustad CM;
XX
DR WPI; 1999-508506/42.
XX
DR N-PSDB; AA209235, AAY09236.
XX
PT New human proteolytic accessory enzyme and its modulators useful
PT for treating disease conditions like inflammation or autoimmune
PT diseases
XX
CC Claim 1; Fig 3; 95pp; English.
XX
CC This invention describes a novel human polynucleotide (I) which encodes a
CC E3 ubiquitin protein ligase, h-E3 UPL (I1). The products of the invention
CC have antiinflammatory, immunosuppressive, neuroprotective, cytosolic,
CC antiarthritic, immunomodulator, antidiabetic, antiviral and cardiac
CC activity. The products of the invention can also be used for treating
CC patients with disorders mediated by the biological and/or pharmacological
CC activity of h-E3 UPL. The nucleic acid sequences encoding E3 UPL are used
CC in expression systems as assay for agonists and antagonists for the E3
CC UPL protein. The E3 UPL protein is used in screening assays to identify
CC blockers and antagonists. They are also used in gene therapy. Specific
CC modulation of biological and/or pharmacological activity of novel h-E3
CC UPL via administration of a modulator or heterologous expression, is used
CC for treating physiological conditions like inflammation, autoimmune
CC diseases, neurological disease, apoptosis, endothelial cell physiology
CC (e.g., proliferation, differentiation), peripheral vascular disease,
CC angiogenesis, cancer, hematopoietic disorders, arthritis, cachexia,
CC leukemia, pulmonary disorders, diabetes and viral infection. The
CC nucleotide sequences which encode h-E3 UPL may also be employed in
CC analysis to map chromosomal location e.g., screening for functional
CC association with disease markers. They are also used as screening tools
```



CC in the identification of appropriate human subjects and patients for  
CC therapeutic clinical trials. The sequences can also be used to detect the  
CC presence of the mRNA transcripts in a patient or to monitor the  
CC modulation of transcripts during treatment. This sequence represents the  
CC human E3 ubiquitin protein ligase protein described in the method of the  
CC invention.  
XX

SQ Sequence 852 AA;

Query Match 18.8%; Score 67.5; DB 20; Length 852;

Best Local Similarity 31.1%; Pred. No. 14;

Matches 23; Conservative 12; Mismatches 20; Indels 19; Gaps 3;

QY 2 SPTLTQSTNTHQTSSSSSDG-----GLFRSRPAH-----SLPPGEDGRVEPYVDF 48

Db 232 SSTGSLPTNTNTNTSGATGGLIPLTISGSGPRPLNPTVQAFLPPGWGEQRVD----- 286

QY 49 AEFYRLMSVDHGEQ 62

Db 287 -QHGRVYVVDHVEK 299

Search completed: December 10, 2002, 07:30:42  
Job time : 87 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 07:27:32 ; Search time 52 Seconds

(without alignment)  
125.714 Million cell updates/sec

Title: US-09-830-144-4\_COPY\_437\_504

Perfect score: 359

Sequence: 1 QSPRLTQSTHTTQSSSSS.....AEFRLMSVDHGEQSVTAP 68

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63.5	17.7	243	2 D84792	probable protein t
2	62	17.3	541	2 S51799	nucleoporin NUP57
3	62	17.3	679	2 H95036	glycosyl hydrolase
4	62	17.3	737	2 D97907	alpha-xylosidase (
5	62	17.3	4957	2 T03455	ALR protein - huma
6	62	17.3	5262	2 T03454	ALR protein - huma
7	61.5	17.1	238	2 T52505	hypothetical prote
8	61.5	17.1	737	2 T52505	hypothetical prote
9	61	17.0	267	2 S74415	S-protein secretio
10	61	17.0	665	2 E82506	hypothetical prote
11	61	17.0	909	2 T06246	proteinase II YCA0
12	61	17.0	951	2 E88042	aspartate kinase (
13	61	17.0	1511	2 A53151	protein F6D12.6 (
14	60	16.7	1260	2 S60896	pleiotropic drug r
15	60	16.7	1576	2 AE0249	agglutinin-like pr
16	59.5	16.6	1369	2 T32338	probable hemolysin
17	59	16.4	173	2 G86861	hypothetical prote
18	59	16.4	329	2 S61884	carbonate dehydrat
19	59	16.4	330	1 S61883	carbonate dehydrat
20	59	16.4	330	1 S61882	carbonate dehydrat
21	59	16.4	429	1 T38146	dihydrofolate redu
22	59	16.4	651	2 UC7705	death receptor-6 -
23	58.5	16.3	378	2 H72679	hypothetical prote
24	58.5	16.3	378	2 A84151	hypothetical prote
25	58.5	16.3	608	2 T53269	prolactin receptor
26	58.5	16.3	1093	2 T38533	AFV protein - hum
27	58.5	16.3	1099	2 T18257	phospholipase C -
28	58.5	16.3	2897	2 B48666	cell proliferation
29	58.5	16.3	3256	2 A48666	cell proliferation

30	58	16.2	142	2 D69891	ynek protein - Bac
31	58	16.2	277	2 D69158	sensory transducti
32	58	16.2	532	2 B35621	spore germination
33	57.5	16.0	172	2 T36107	probable serine/ar
34	57.5	16.0	232	1 A25108	homeotic protein/H
35	57.5	16.0	366	2 A86392	hypothetical prote
36	57.5	16.0	394	2 H65010	hypothetical prote
37	57.5	16.0	396	2 A57090	CSA protein - huma
38	57.5	16.0	629	2 B83107	chemotactic transd
39	57.5	16.0	686	2 UC5708	villin-like protei
40	57	15.9	762	2 S67765	probable membrane
41	57	15.9	778	2 H84678	hypothetical prote
42	56.5	15.7	276	2 T06331	photosystem II 22k
43	56.5	15.7	283	2 H83860	pantothenate synth
44	56.5	15.7	343	2 E83673	soybitol dehydroge
45	56.5	15.7	457	2 B64790	ycbB protein - Esc

## ALIGNMENTS

## RESULT 1

D84792  
probable protein translocase (imported) - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C/Accession: D84792

R/Lin: X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; I

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Umayam, L.; Tallon, L.

guss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J

Nature 402, 761-768, 1999

A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A/Reference number: A84420; MID:20083487; PMID:10617197

A/Accession: D84792

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-243 <STO>

A/Cross-References: GB:A8002093; NID:94056494; PIDN:AC98060.1; GSPDB:GN00139

C/Genetics:

A/Map position: 2

Query Match 17.7%; Score: 63.5; DB 2; Length 243;  
Best Local Similarity 39.0%; Prod. No. 7.9;  
Matches 16; Conservative 6; Mismatches 18; Indels 1; Gaps 1;

QY 8 QSTNHTQSSSSSDGIFRSPAHSLPGEEDGRVPEYDF 48  
Db 193 QNONTASSSSSWFGGLF-DKKKEVQPSSEKTEVLESF 232

RESULT 2  
S51799  
nucleoporin NUP57 - yeast (Saccharomyces cerevisiae)

N/Alternate names: protein G6320; protein YGR119C

A;Cross-references: EMBL:X81155  
 R:Van Dyck, L.; Skala, J.; de Wergifosse, P.; Purnelle, B.; Talla, E.; Nawrocki, A.; Del  
 submitted to the Protein Sequence Database, May 1996  
 A;Reference number: S64428  
 A;Accession: S64428  
 A:Molecule type: DNA  
 A;Residues: 1-353 <VAN>  
 A;Cross-references: EMBL:272904; MIPS:YGR119C  
 A;Experimental source: strain S288C  
 R:Hansen, M.; Albers, M.; Backes, U.; Coblentz, A.; Leuther, H.; Neu, R.; Schreier, A.; Sch  
 submitted to the Protein Sequence Database, May 1996  
 A;Reference number: S64417  
 A;Accession: S64417  
 A:Molecule type: DNA  
 A;Residues: 300-541 <HAN>  
 A;Cross-references: EMBL:272904; MIPS:YGR119C  
 A;Experimental source: strain S288C  
 R:van Dyck, L.; Goffeau, A.  
 submitted to the EMBL Data Library, December 1994  
 A;Description: Genes for an asn synthase, a GLFG-motif nucleoporin and a putative homeob  
 e new ORFs, remnants of Ty and three tRNA genes.  
 A;Reference number: S55976  
 A;Accession: S55976  
 A:Molecule type: DNA  
 A;Residues: 1-353 <VAV>  
 A;Cross-references: EMBL:X83099  
 R:Hansen, M.; Albers, M.; Backes, U.; Coblentz, A.; Leuther, H.; Neu, R.; Schreier, A.; Sch  
 Yeast 12, 1273-1277, 1996  
 A;Title: The sequence of a 23.4 kb segment on the right arm of chromosome VII from Sacch  
 A;Reference number: S72192  
 A;Accession: S72192  
 A:Molecule type: DNA  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Residues: 1-541 <HAW>  
 A;Cross-references: EMBL:272904; NID:G1323192; PIDN:CAA97129.1; PID:G1323193  
 A;Note: The nucleotide sequence was submitted to the EMBL Data Library, May 1996  
 C;Genetics:  
 A;Gene: SGD:NUP57  
 A;Cross-references: SGD:S0003351; MIPS:YGR119C  
 A;Map position: 7R

Query Match 17.3%; Score 62; DB 2; Length 541;  
 Best Local Similarity 50.0%; Pred. No. 30;  
 Matches 12; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 8 QSTNTHTQSSSSSDGGLFRSRPA 31  
 DB 60 QATNTGSGNQSSGTGGGLFCNRP 83

RESULT 3  
 H95036  
 glycosyl hydrolase, family 31 SP0312 [imported] - Streptococcus pneumoniae (strain TIGR4  
 C;Species: Streptococcus pneumoniae  
 C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
 C;Accession: H95036  
 R:Tetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.K.; Radune, D.; Holtzapf  
 nson, T.; Hickey, E.K.; Holt, I.E.  
 Science 293, 498-506, 2001  
 A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
 A;Reference number: A95000; MUID:21357209; PMID:11463916  
 A;Accession: H95036  
 A;Status: preliminary  
 A:Molecule type: DNA  
 A;Residues: 1-679 <KOR>  
 A;Cross-references: GB:AB005672; PIDN:AAK74489.1; PID:G14971785; GSPDB:GN00164; TIGR:SP4  
 A;Experimental source: strain TIGR4  
 C;Genetics:  
 A;Gene: SP0312

Query Match 17.3%; Score 62; DB 2; Length 679;

Best Local Similarity 32.8%; Pred. No. 39;  
 Matches 21; Conservative 4; Mismatches 23; Indels 16; Gaps 3;

QY 13 HTQSSSSSDGGLFRSRPA---HSLPPGEGD-----RVEPYVDFAEF---YRLWS 56  
 DB 289 HYQDSCKNAEGGLILSRVAGPGSHRYPVGFSGDTIISWNLRFQPYFTATASNIGYSWWS 348

QY 57 VDHG 60  
 DB 349 HDIG 352

RESULT 4  
 D97907  
 alpha-xylosidase (EC 3.2.1.1-) [imported] - Streptococcus pneumoniae (strain R6)  
 C;Species: Streptococcus pneumoniae  
 C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
 C;Accession: D97907  
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E  
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
 Y, P.; Sun, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaaskunas, S.R.;  
 A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
 A;Reference number: A97872; MUID:21429245; PMID:11544234  
 A;Accession: D97907  
 A;Status: preliminary  
 A:Molecule type: DNA  
 A;Residues: 1-737 <KUR>  
 A;Cross-references: GB:AB007317; PIDN:AAK99088.1; PID:G15457836; GSPDB:GN00174  
 C;Genetics:  
 A;Gene: xylS  
 C;Keywords: glycosidase; hydrolase

Query Match 17.3%; Score 62; DB 2; Length 737;  
 Best Local Similarity 32.8%; Pred. No. 42;  
 Matches 21; Conservative 4; Mismatches 23; Indels 16; Gaps 3;

QY 13 HTQSSSSSDGGLFRSRPA---HSLPPGEGD-----RVEPYVDFAEF---YRLWS 56  
 DB 347 HYQDSCKNAEGGLILSRVAGPGSHRYPVGFSGDTIISWNLRFQPYFTATASNIGYSWWS 406

QY 57 VDHG 60  
 DB 407 HDIG 410

RESULT 5  
 T03455  
 ALR protein - human  
 C;Species: Homo sapiens (man)  
 C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
 C;Accession: T03455  
 R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,  
 Oncogene 15, 549-560, 1997  
 A;Title: Structure and expression pattern of human ALR, a novel gene with strong homology  
 A;Reference number: Z14954; MUID:97388474; PMID:9247308  
 A;Accession: T03455  
 A;Status: preliminary; translated from GB/EMBL/DDBU  
 A:Molecule type: mRNA  
 A;Residues: 1-4957 <PRA>  
 A;Cross-references: EMBL:AF010404; NID:G2358286; PIDN:AAK51735.1; PID:G2358287  
 C;Genetics:  
 A;Gene: ALR  
 A;Map position: 12  
 A;Superfamily: human ALR protein  
 C;Keywords: alternative splicing

Query Match 17.3%; Score 62; DB 2; Length 4957;  
 Best Local Similarity 33.3%; Pred. No. 3.9e+02;  
 Matches 26; Conservative 8; Mismatches 22; Indels 22; Gaps 6;

QY 1 QSTLTTLQSTNTHQTSS---SSSSDG---GLFRSRPAHSL-PPGEGRVEPYVDFAEFY 52

```

Db 1996 OSTNYTAVTGNFHPHSGPLCPSSSGTGESYGLSPLRPSVLPFPAPDGL-PY----- 2047
OY 53 RLMSVDHG--EQSVVTP 68
Db 2048 ----LSHGASQSGITSP 2061

```

## RESULT 6

```

T03454
ALR protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03454
R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
A:Reference number: Z14954; MID:97388474; PMID:9247308
A:Accession: T03454
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5262 <PRA>
A:Cross-references: EMBL:AF010403; NID:92358284; PIDN:AAC51734.1; PID:92358285
C:Genetics:
A:Gene: ALR
A:Map position: 12
C:Superfamily: human ALR protein
C:Keywords: alternative splicing

```

```

Query Match 17.3%; Score 62; DB 2; Length 5262;
Best Local Similarity 33.3%; Pred. No. 41e+02;
Matches 26; Conservative 8; Mismatches 22; Indels 22; Gaps 6;

OY 1 QSPFLTLQSTNTHHTQSS-----SSSDG---GLFRSRPAHSL-PPGEDGRVEPYVDFAFYF 52
Db 2301 QSTNYTAVTGNFHPHSGPLCPSSSGTGESYGLSPLRPSVLPFPAPDGL-PY----- 2352

OY 53 RLMSVDHG--EQSVVTP 68
Db 2353 ----LSHGASQSGITSP 2366

```

## RESULT 7

```

T52505
hypothetical protein B2023.10 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T52505
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, September 2000
A:Reference number: Z26053
A:Accession: T52505
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-238 <SCH>
A:Cross-references: EMBL:AF442164; GSPDB:GN00116; NCSP:B2023.10
A:Experimental source: BAC clone B2023; strain OR74A
C:Genetics:
A:Gene: NCSP:B2023.10
A:Map position: 6
A:Introns: 188/2

```

```

Query Match 17.1%; Score 61.5; DB 2; Length 238;
Best Local Similarity 31.6%; Pred. No. 13;
Matches 24; Conservative 8; Mismatches 27; Indels 17; Gaps 3;

OY 2 SPTTLTQSTNTHHTQSSSSSDGGLFR-----SRPAHSLPPGEDG-----RVEPYVDF 48
Db 24 SPT-----ATNTSSSSSSSSSSSAFGLPRPISYPTALIGPRRGGPKTPCKNGPVLDQ 79
OY 49 AEFRLMSVDHGEQSV 64
Db 80 TSAGSPYSVDCKQISI 95

```

## RESULT 8

```

139547
S-protein secretion D - Aeromonas hydrophila
C:Species: Aeromonas hydrophila
C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 08-Oct-1999
C:Accession: I39547
R:Thomas, S.R.; Trust, T.J.
J. Bacteriol. 177, 3932-3939, 1995
A:Title: A specific Pseudomonas sp. is required for the secretion of paracrystalline surface
A:Reference number: A57354; MID:95332195; PMID:7608063
A:Accession: I39547
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-737 <RES>
A:Cross-references: GB:I41682; NID:950300; PIDN:AAA79322.1; PID:91019923
C:Genetics:
A:Gene: spsd

```

```

Query Match 17.1%; Score 61.5; DB 2; Length 737;
Best Local Similarity 37.8%; Pred. No. 48;
Matches 17; Conservative 3; Mismatches 20; Indels 5; Gaps 1;

OY 3 PTLTQSTNTHHTQSSSSSDGGLFRSRPAHSL-----LPPGEDGRV 42
Db 385 PATGVNSTVTATGEGSGAGSLTAAAPATAPTSTAPRGEQGSV 429

```

## RESULT 9

```

S74415
hypothetical protein s110688 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S74415
R:Kaneko, T.; Sato, S.; Kori, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;
O.K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A:Reference number: S74322; MID:97061201; PMID:8905231
A:Accession: S74415
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-267 <KAN>
A:Cross-references: EMBL:D64001; GB:AB001339; NID:91001102; PIDN:BA10333.1; PID:d101098
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

```

```

Query Match 17.0%; Score 61; DB 2; Length 267;
Best Local Similarity 29.9%; Pred. No. 17;
Matches 20; Conservative 12; Mismatches 21; Indels 14; Gaps 3;

OY 1 QSPFT--LTLQSTNTHHTQSSSSSDGGLFRSR--PAHSLPPGEDGRVEPYVDFAFYFLW 55
Db 153 QEPNKNFTLTLRATQAKQATGSLIVGASASALLVLRDVPYIKQKV-----YQLW 203

OY 56 SYDHGEQ 62
Db 204 AVIDGKK 210

```

## RESULT 10

```

B82506
protease II VCA0063 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82506
R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bess, S.; Qin, H.; Dragoti, I.; Sellers, P.
J. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

```



F:560-576/Domain: transmembrane #status predicted <TM2>  
F:612-628/Domain: transmembrane #status predicted <TM3>  
F:633-649/Domain: transmembrane #status predicted <TM4>  
F:663-679/Domain: transmembrane #status predicted <TM5>  
F:777-794/Domain: transmembrane #status predicted <TM6>  
F:888-1087/Domain: transmembrane #status predicted <TM7>  
F:905-912/Region: nucleotide-binding motif A (P-loop)  
F:1036-1252/Domain: transmembrane #status predicted <TM7>  
F:1270-1295/Domain: transmembrane #status predicted <TM8>  
F:1330-1346/Domain: transmembrane #status predicted <TM9>  
F:1358-1374/Domain: transmembrane #status predicted <TM10>  
F:1389-1405/Domain: transmembrane #status predicted <TM11>  
F:1478-1495/Domain: transmembrane #status predicted <TM12>

Query Match 17.0%; Score 61; DB 2; Length 1511;  
Best Local Similarity 38.8%; Pred. No. 1.3e+02;  
Matches 19; Conservative 6; Mismatches 16; Indels 8; Gaps 3;

QY 4 TLTLQSTNTHQTSSSSSDGGLFRSPAHSLP--PGEDGRVDP 44  
DB 49 TLTAQSMQNSTQSAAPKSDAQSIFFSGVGVNPIFSDPEAPGYDPKLDLP 97

## RESULT 14

S60896 agglutinin-like protein - yeast (Candida albicans)

C/Species: Candida albicans  
C/Date: 27-Apr-1996 #sequence\_revision 13-Mar-1997 #text\_change 17-Mar-2000  
C/Accession: S60896

R/Hoyer, L.L.; Scherer, S.; Shatzman, A.R.; Livi, G.P.

Mol. Microbiol. 15, 39-54, 1995

A/Title: Candida albicans ALS1: domains related to a Saccharomyces cerevisiae sexual agglutinin-like protein

A/Reference number: S60896; MUID:95272392; PMID:7752895

A/Accession: S60896

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1260 <HOY>

A/Cross-References: EMBL:L25902; NID:G704426; PIDN:AAC41649.1; PID:G704427

C/Superfamily: Yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

Query Match 16.7%; Score 60; DB 2; Length 1260;  
Best Local Similarity 24.2%; Pred. No. 1.3e+02;  
Matches 23; Conservative 9; Mismatches 35; Indels 28; Gaps 3;

QY 2 SPTLT-----QSTNTHQTSSSSSDGGLFRSPAHSL-----PP 36  
DB 651 NPTVTTEYSOSYATTTTITAPPGEDTVLIREPPNHTVTTEYSOSYATTTVTAP 710

QY 37 GEDGRV---EPYVDFAEFYRLMSVDHGOSVVTAP 68  
DB 711 GEDTVLIREPPNHTVTTEYSOSYATTTVTAP 745

## RESULT 15

AE0249 probable hemolysin YPO2045 [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis

C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001

C/Accession: AE0249

R/Parhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Farrage, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;

11, M.; Rutledge, K.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001

A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AE0249

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1576 <YRS>

A/Cross-References: GB:AL590842; PIDN:CAC90857.1; PID:G15980056; GSPDB:GN00175

C/Genetics: YPO2045

Query Match 16.7%; Score 60; DB 2; Length 1576;  
Best Local Similarity 25.4%; Pred. No. 1.7e+02;  
Matches 17; Conservative 10; Mismatches 20; Indels 20; Gaps 3;

QY 5 TLTLQSTNTHQTSSSSSDGGLFRSPAHSLPPEGEDGRVDPYDFA----EF---YRLW 55  
DB 506 IDQSKNTHIGAKIASGEL-----SIDAKGVYIGLASMLTSEFKDIDYMW 554

QY 56 SYDHGEQ 62  
DB 555 GGAHGE 561

Search completed: December 10, 2002, 07:34:35  
Job time : 56 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 06:49:56 ; Search time 42 Seconds  
(without alignments)  
67,152 Million cell updates/sec

Title: US-09-830-144-4\_COPY\_437\_504

Perfect score: 359  
Sequence: 1 QSPFTLLQSTNTHTQSSSSS.....AEFYRLMSVDHGEOGVTPAP 68

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	359	100.0	504	1 TAB1_HUMAN	Q15750 homo sapien
2	63.5	17.7	243	1 IM17_ARATH	O99P35 arabidopsis
3	63.5	17.7	824	1 AD08_HUMAN	P78325 homo sapien
4	62	17.3	541	1 NU07_YEAST	P48837 saccharomyc
5	61	17.0	1119	1 AL03_CANAL	O74623 candida alb
6	61	17.0	1511	1 PD05_YEAST	P33302 saccharomyc
7	60	16.7	226	1 OAZ_MOUSE	P54369 mus musculu
8	60	16.7	1260	1 AL01_CANAL	P46590 candida alb
9	59	16.4	328	1 CAH1_PLAAR	P46281 flaveria pr
10	59	16.4	330	1 CAH1_PLAAR	P46512 flaveria pr
11	59	16.4	330	1 CAH1_PLAAR	P46511 flaveria pr
12	59	16.4	399	1 BR03_SHEEP	O97967 ovis aries
13	58.5	16.3	608	1 PRLR_MOUSE	O08501 mus musculu
14	58.5	16.3	1093	1 AF17_HUMAN	P55198 homo sapien
15	58.5	16.3	1099	1 PL01_CANAL	O13433 candida alb
16	58.5	16.3	3256	1 K167_HUMAN	P46013 homo sapien
17	58	16.2	142	1 YNEK_BACSU	P45711 bacillus su
18	58	16.2	276	1 CAPB_DROME	P46603 drosophila
19	58	16.2	532	1 SP07_DICDI	P26598 dictyostel
20	57.5	16.0	232	1 HXB4_XENLA	P09070 xenopus lae
21	57.5	16.0	396	1 CSA_HUMAN	O13216 homo sapien
22	57.5	16.0	686	1 V1LE_HUMAN	O13195 homo sapien
23	57	15.9	937	1 NU08_HUMAN	P59248 homo sapien
24	57	15.9	1377	1 NE01_RAT	P97603 rattus norv
25	57	15.9	1461	1 NE01_HUMAN	O92859 homo sapien
26	56.5	15.7	276	1 PSB5_LYCES	P54773 lycopersico
27	56.5	15.7	283	1 PANC_BACHD	O94086 bacillus ha
28	56.5	15.7	437	1 DUSC_ECOLI	O92901 bacillus ha
29	56.5	15.7	453	1 HUSC_ECOLI	P77211 escherichia
30	56.5	15.7	459	1 MCE1_YEAST	O01195 saccharomyc
31	56.5	15.7	463	1 PL0B_CARTI	O04213 carthamus t
32	56.5	15.7	922	1 NR01_RAT	O90019 rattus norv
33	56.5	15.7	923	1 NR01_HUMAN	O14786 homo sapien

34	56.5	15.7	1858	1 P3K2_DICDI	P54674 dictyostel
35	56	15.6	297	1 PRTA_SPRGR	P00776 streptomyce
36	56	15.6	315	1 SPY2_HUMAN	O43597 homo sapien
37	56	15.6	325	1 GELP_SOYBN	O39836 glycine max
38	56	15.6	405	1 Y574_HUMAN	O60320 homo sapien
39	56	15.6	427	1 TR16_HUMAN	P08138 homo sapien
40	56	15.6	509	1 RUNT_DROME	P22814 drosophila
41	56	15.6	639	1 BARI_SCHCO	O92275 schizophy11
42	56	15.6	652	1 FX01_MOUSE	O91160 mus musculu
43	56	15.6	743	1 TLE2_HUMAN	O04725 homo sapien
44	56	15.6	776	1 AD28_MACFA	O9X516 macaca fasc
45	56	15.6	889	1 GYRA_STRAO	P20831 staphylococ

## ALIGNMENTS

RESULT 1	TAB1_HUMAN	STANDARD:	PRT:	504 AA.
ID	TAB1_HUMAN			
AC	Q15750;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	TAK1-binding protein 1.			
GN	TAB1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=96215294; PubMed=8638164;			
RA	Shibuya H., Yamaguchi K., Shirakabe K., Torigawa A., Gotch Y.,			
RA	Ueno N., Irie K., Nishida E., Matsumoto K.;			
RT	"TAB1: an activator of the TAK1 MAPKK in TGF-beta signal			
RT	transduction.";			
RL	Science 272:1179-1182(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20057165; PubMed=10591208;			
RA	Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,			
RA	Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,			
RA	Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,			
RA	Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,			
RA	Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,			
RA	Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,			
RA	Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,			
RA	Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,			
RA	Hall R., Hall J., Hall T., Hall M., Heathcote R.W., Ho S., Holmes S.,			
RA	Hunt S.E., Jones M.C., Kersey J., Kimberley A.M., King A.,			
RA	Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,			
RA	Marvin I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann B.J.,			
RA	McClay J., McLaren S., McMuray A.A., Milne S.A., Morrison B.J.,			
RA	Odeh C.N., Pavlic R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,			
RA	Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,			
RA	Scott C.E., Senta H.K., Skuce C.D., Smalley S., Smith M.L.,			
RA	Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,			
RA	Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,			
RA	Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,			
RA	Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,			
RA	Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,			
RA	Shirahata A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyma S.,			
RA	Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,			
RA	Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,			
RA	Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,			
RA	Pan S., Qi S., Qian Y., Ray L., Ren Q., Shauli S., Sloan D., Song L.,			
RA	Phang O., Wang Y., Wang Z., White J., Williamson D., Wu H., Yao Z.,			
RA	Zhan M., Zhang G., Chissole S., Murray J., Miller N., Mink P.,			
RA	Fulton R., Johnson D., Bemis G., Bentley D., Bredshaw H., Bourne S.,			

```
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Lattelle P., Layman D., Ozerky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Neilson J.,
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA Kim U.J., Shituya H., Simon M.I., Dumanaki J.P., Peyrard M., Kedra D.,
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tiliahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
CC -!- FUNCTION: MAY BE AN IMPORTANT SIGNALING INTERMEDIATE BETWEEN TGFB
CC RECEPTORS AND MAP3K7/TAK1.
CC -!- SUBUNIT: INTERACTS WITH MAP3K7 AND WITH BIRC7.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: CONTAINS 1 PP2C-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U49928; AAC12660.1; -
DR EMBL; Z83845; CAB55304.1; -
DR Genbank; HGNC:18157; MAP3K7IP1.
DR MIM; 602615; -
DR InterPro; IPR001932; PP2C-like.
DR Pfam; PF00481; PP2C.1.
DR SMART; SM00332; PP2CC; 1.
FT DOMAIN 64 368 PP2C-LIKE.
FT DOMAIN 452 457 POLY-SER.
SQ SEQUENCE 504 AA; 54644 MW; A45743288718903A CRC64;

Query Match 100.0%; Score 359; DB 1; Length 504;
Best Local Similarity 100.0%; Pred. No. 6.9e-34;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QSTPLTQSTNTHTQSSSSDGGFLFRSRPAHSLPPGEGDGRVEPYVDFAEYRLWSVDHG 60
DB 437 QSTPLTQSTNTHTQSSSSDGGFLFRSRPAHSLPPGEGDGRVEPYVDFAEYRLWSVDHG 496

QY 61 EQSVVTPAP 68
DB 497 EQSVVTPAP 504

RESULT 2
ID IM17 ARATH STANDARD; PRT; 243 AA.
AC Q9SP15; Q9ZUS5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitochondrial import inner membrane translocase subunit TIM17.
OS Arabidopsis thaliana (Mouse-ear cress).
GN Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Miernyk J.A., Coop N.E.;
RT "A component of the Arabidopsis thaliana mitochondrial inner membrane
RT protein translocase, atTIM17."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271556; PubMed=9126482;
RA Yoshiyama K., Higuchi Y., Kataoka M., Matsuura K., Yamamoto S.;
```

```
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Unayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Niernan W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768(1999).
CC -!- FUNCTION: COMPONENT OF THE PREPROTEIN IMPORT MACHINERY OF THE
CC MITOCHONDRIAL INNER MEMBRANE. INTEGRAL PART OF A PROTEIN-
CC CONDUCTING CHANNEL OF THE MITOCHONDRIAL INNER MEMBRANE (BY
CC SIMILARITY).
CC -!- SUBUNIT: FORMS PART OF THE TIM23 RECEPTOR COMPLEX THAT CONSISTS OF
CC AT LEAST 3 DIFFERENT PROTEINS; TIM17, TIM23 AND TIM44 (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane (BY similarity).
CC -!- SIMILARITY: BELONGS TO THE TIM17/TIM22/TIM23 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF186847; AAF03749.1; -
DR EMBL; AC005896; AAC98060.1; -
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003397; Tim17.
DR Pfam; PF02466; Tim17; 1.
KW Transport; Protein transport; Translocation; Mitochondrion;
KW Inner membrane; Transmembrane
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 83 83 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT CONFLICT 116 116 A -> T (IN REF. 1).
FT CONFLICT 219 219 Q -> H (IN REF. 1).
SQ SEQUENCE 243 AA; 25571 MW; 199285297F58BD51 CRC64;

Query Match 17.7%; Score 63.5; DB 1; Length 243;
Best Local Similarity 39.0%; Pred. No. 3.1;
Matches 16; Conservative 6; Mismatches 18; Indels 1; Gaps 1;

QY 8 QSTNTHTQSSSSDGGFLFRSRPAHSLPPGEGDGRVEPYVDV 48
DB 193 QNQTASSSSSSWFGGLF-DKKKEVQPGSEKTEVLESF 232

RESULT 3
AD08_HUMAN
ID AD08_HUMAN STANDARD; PRT; 824 AA.
AC F78325;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ADAM 8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE domain 8) (Cell surface antigen MS2) (CD156a antigen) (CD156).
GN ADAM8 OR MS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=97271556; PubMed=9126482;
RA Yoshiyama K., Higuchi Y., Kataoka M., Matsuura K., Yamamoto S.;
```

```

RT "CD156 (human ADAM8) : expression, primary amino acid sequence, and
RL gene location."
CC -1- FUNCTION: POSSIBLE INVOLVEMENT IN EXTRAVASATION OF LEUCOCYTES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSED ON NEUTROPHILS AND MONOCYTES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.
CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD156 entry;
WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd156.htm"
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D26579; BAA05626.1; -.
CC HSSP; P18619; 1FVL.
CC MEROPS; M12.208; -.
CC Genew; HGNC:215; ADAM8.
CC
CC MTM; 602267; -.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002870; EGF M12B propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000130; Zn_MTPeptidase.
CC Pfam; PF00200; disintegrin; 1.
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC ProDom; PD000664; Disintegrin; 1.
CC SMART; SM00050; Disin; 1.
CC SMART; SM00181; EGF; 1.
CC PROSITE; PS50215; ADAM_MERPRO; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE; PS50214; DISINTEGRIN_2; 1.
CC PROSITE; PS01186; EGF_2; UNKNOWN_1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC HydroLase; Metalloprotease; Zinc; Signal; Glycoprotein;
KM Transmembrane; Antigen.
FT CHAIN 1 16
FT SIGNAL 1 16
FT DOMAIN 17 824 ADAM 8.
FT TRANSMEM 656 676 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 677 824 POTENTIAL.
FT DOMAIN 200 400 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 408 494 METALLOPROTEASE.
FT METAL 334 334 ZINC (CATALYTIC) (PROBABLE).
FT ACT_SITE 335 335 BY SIMILARITY.
FT METAL 338 338 ZINC (CATALYTIC) (PROBABLE).
FT METAL 344 344 ZINC (CATALYTIC) (PROBABLE).
FT DISULFID 310 395 BY SIMILARITY.
FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 436 436 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 824 AA; 88673 MW; 5DF8E05F30DF479E CRC64;

Query Match 17.7%; Score 63.5; DB 1; Length 824;
Best Local Similarity 38.6%; Pred. No. 13;
Matches 27; Conservative 2; Mismatches 30; Indels 11; Gaps 3;
OY 7 LOSTNTHTOSSSSSSD-----GLFRSRPAHSILPPGSDGKVEVY--VDPAEFYRLWS-- 56
DB 161 LQTAGTCGVSDSLGSLGRTAAVFRPRGDSLPSRETRVELLVVVVDNAEFGMLGSEA 220
OY 57 -VDHGEQSVV 65
DB 221 AVRRHVELEV 230

```

```

RESULT 4
ID NUS7_YEAST STANDARD; PRT; 541 AA.
AC P48837;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleoporin NUP57 (Nuclear pore protein NUP57).
GN NUP57 OR YGR119C OR G6320.
OS Saccharomyces cerevisiae (baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95129554; PubMed=7828598;
RA Grandi P., Schlaich N., Tekotte H., Hurt E.C.;
RT "Functional interaction of Nup57p with a core nucleoporin complex
RL consisting of Nup1p, Nup49p and a novel protein Nup57p.";
RL EMBO J. 14:76-87(1995).
RN [2]
RP SEQUENCE OF 1-354 FROM N.A.
RC STRAIN=S288c / F71679;
RX MEDLINE=97197982; PubMed=9046098;
RA Van Dyck L., Tetteijn H., Purnelle B., Goffeau A.;
RT "An 18.3 kb DNA fragment from yeast chromosome VII carries four
RT unknown open reading frames, the gene for an Asn synthase, remnants
RL of Ty and three tRNA genes.";
RL Yeast 13:171-176(1997).
RN [3]
RP SEQUENCE OF 243-541 FROM N.A.
RX MEDLINE=97061913; PubMed=8905931;
RA Hansen M., Albers M., Backes U., Coblentz A., Leuther H., Neu R.;
RA Schreier A., Schaefer B., Zimmermann M., Wolf K.;
RT "The sequence of a 23.4 kb segment on the right arm of chromosome VII
RT from Saccharomyces cerevisiae reveals CLB6, SPT6, RPS28A and NUP57
RL genes, a Ty3 element and 11 new open reading frames.";
RL Yeast 12:1273-1277(1996).
CC -1- FUNCTION: PART OF THE NUCLEOPORIN COMPLEX; REQUIRED FOR PROTEIN
CC TRANSPORT IN THE NUCLEUS.
CC -1- SUBUNIT: INTERACTS WITH NSP1, NUP49 AND NUP96.
CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X81155; CAA57053.1; -.
CC EMBL; X83099; CAA58153.1; -.
CC EMBL; 272904; CAA97129.1; -.
CC EMBL; 272905; CAA97131.1; -.
CC SCD; S0003351; NUP57.
KM Nuclear protein; Transport; Coiled coil; Repeat.
FT DOMAIN 76 223 9 X 4 AA REPEATS OF G-L-F-G.
FT DOMAIN 26 31 POLY-ASN.
FT DOMAIN 127 130 POLY-THR.
FT DOMAIN 217 220 POLY-GLY.
FT DOMAIN 258 266 POLY-GLN.
FT DOMAIN 277 280 POLY-GLN.
FT DOMAIN 398 425 COILED COIL (POTENTIAL).
SQ SEQUENCE 541 AA; 57498 MW; B292ADPF7B1D7E83C CRC64;

Query Match 17.3%; Score 62; DB 1; Length 541;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 12; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
OY 8 QSTNTHTOSSSSSDGGLFRSRPA 31

```

```

Db 60 QATNTFGSNQSSSTGGGLFGNKPA 83
RESULT 5
ALS3_CANAL STANDARD; PRT; 1119 AA.
AC 074623;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Agglutinin-like protein 3 precursor.
GN ALS3.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RX MEDLINE=98309840; PubMed=9644209;
RA Hoyer L.L., Payne T.L., Bell M., Myers A.M., Scherer S.;
RT "Candida albicans ALS3 and insights into the nature of the ALS gene family."
RL Curr. Genet. 33:451-459(1998).
CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U87956; AAC39486.1; -.
KW Cell adhesion; Glycoprotein; Repeat; Signal.
FT SIGNAL 1 17
FT CHAIN 18 1119
FT DOMAIN 433 792
FT REPEAT 433 468
FT REPEAT 469 504
FT REPEAT 505 540
FT REPEAT 541 576
FT REPEAT 577 612
FT REPEAT 613 648
FT REPEAT 649 684
FT REPEAT 685 720
FT REPEAT 721 756
FT REPEAT 757 792
FT DOMAIN 399 404
FT DOMAIN 450 455
FT DOMAIN 557 563
FT DOMAIN 593 597
FT DOMAIN 630 635
FT DOMAIN 666 671
FT DOMAIN 702 707
FT DOMAIN 738 743
FT DOMAIN 774 777
FT DOMAIN 1044 1047
FT CARBOHYD 471 471
FT CARBOHYD 543 543
FT CARBOHYD 579 579
FT CARBOHYD 651 651
FT CARBOHYD 687 687
FT CARBOHYD 723 723
FT CARBOHYD 759 759
FT CARBOHYD 845 845
FT CARBOHYD 987 987
FT CARBOHYD 1050 1050
FT CARBOHYD 1061 1061
FT SEQUENCE 1119 AA; 119927 MW; 6A3PB3FC8C879A71 CRC64;

Query Match 17.0%; Score 61; DB 1; Length 1119;
Best Local Similarity 23.8%; Pred. No. 38;
Matches 24; Conservative 13; Mismatches 28; Indels 36; Gaps 4;

QY 1 QSPFTLTQSTNTHTSSSSSS-----SDGLFRSPAHSL----- 34
Db 720 EPPNHTVTTTTEYMSQSYATTITAPPGETDTVLIREPNNHTVTTTTEYMSQSYATTITII 779
QY 35 -PPGEDGRV-----EPYVDFAEFYRLMSVDHGEQSVVTAP 68
Db 780 APPGETDTVLIREPNNPTVTTEY---WSQSYTTATTVTAP 817

RESULT 6
PDR5_YEAST
ID PDR5_YEAST STANDARD; PRT; 1511 AA.
AC P33302;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Suppressor of toxicity of sporidesmin.
GN PDR5 OR STS1 OR YDR1 OR LEM1 OR YOR153W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB320;
RX MEDLINE=94140838; PubMed=8307980;
RA Bissinger P.H., Kuchler K.;
RT "Molecular cloning and expression of the Saccharomyces cerevisiae
RT STS1 gene product. A yeast ABC transporter conferring mycotoxin
RT resistance."
RL J. Biol. Chem. 269:4180-4186(1994).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=S288c;
RX MEDLINE=94124579; PubMed=8294477;
RA Balzi E., Wang M., Leterme S., van Dyck L., Goffeau A.;
RT "PDR5, a novel yeast multidrug resistance conferring transporter
RT controlled by the transcription regulator PDR1."
RL J. Biol. Chem. 269:2206-2214(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AH22;
RX MEDLINE=95188264; PubMed=7882421;
RA Hirata D., Yano K., Miyahara K., Miyakawa T.;
RT "Saccharomyces cerevisiae YDR1, which encodes a member of the ATP-
RT binding cassette (ABC) superfamily, is required for multidrug
RT resistance."
RL Curr. Genet. 26:285-294(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1678;
RX Tarassov I.A., Winsor B., Martin R.P.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TRANSPORTER, WHOSE PHYSIOLOGICAL FUNCTION IS NOT YET
CC ESTABLISHED. CONFERS RESISTANCE TO THE CHEMICALS CYCLOHEXIMIDE
CC AND SULFOMETHURON METHYL. EXHIBITS NUCLEOSIDE TRIPHOSPHATASE
CC ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. PDR5 SUBFAMILY.
CC -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 61.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL; X74113; CAAS52212.1; -.
DR      EMBL; L19922; AAB53769.1; -.
DR      EMBL; D26548; BAA05547.1; ALT_INIT.
DR      EMBL; U55020; AAC49639.1; -.
DR      EMBL; Z75061; CAA99359.1; -.
DR      PIR; S34702; S34702.
DR      PIR; A49730; A49730.
DR      SGD; S0005679; PDR5.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_Transport.
DR      InterPro; IPR005285; PDR.
DR      Pfam; PF00005; ABC_tran; 2.
DR      ProDom; PD000006; ABC_transport; 2.
DR      SMART; SM00382; AAA; 1.
DR      TIGRFAMs; TIGR00956; 3a01205; 1.
DR      PROSITE; PS00211; ABC_TRANSPORTER; 1.
KM      ATP-binding; Transmembrane; Glycoprotein; Transport.
FT      DOMAIN
FT      1 517
FT      TRANSMEM 518 542 POTENTIAL.
FT      TRANSMEM 559 579 POTENTIAL.
FT      TRANSMEM 612 628 POTENTIAL.
FT      TRANSMEM 632 650 POTENTIAL.
FT      TRANSMEM 666 685 POTENTIAL.
FT      TRANSMEM 775 793 POTENTIAL.
FT      DOMAIN 794 1237 CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 1238 1260 POTENTIAL.
FT      TRANSMEM 1292 1313 POTENTIAL.
FT      TRANSMEM 1325 1349 POTENTIAL.
FT      TRANSMEM 1355 1379 POTENTIAL.
FT      TRANSMEM 1389 1407 POTENTIAL.
FT      TRANSMEM 1477 1499 POTENTIAL.
FT      DOMAIN 1500 1511 CYTOPLASMIC (POTENTIAL).
FT      NP_BIND 905 912 ATP (POTENTIAL).
FT      DOMAIN 784 787 POLY-PHE.
FT      CONFLICT 171 171 N -> L (IN REF. 3).
FT      CONFLICT 190 190 V -> I (IN REF. 3).
FT      CONFLICT 214 214 D -> T (IN REF. 3).
FT      CONFLICT 308 308 G -> V (IN REF. 3).
FT      CONFLICT 340 345 MISSING (IN REF. 3).
FT      CONFLICT 476 476 R -> H (IN REF. 3).
FT      CONFLICT 648 648 MISSING (IN REF. 3).
FT      CONFLICT 770 770 D -> H (IN REF. 3).
SQ      SEQUENCE 1511 AA; 170437 MW; 4540DC0BF04744BA CRC64;

Query Match      17.0%; Score 61; DB 1; Length 1511;
Best Local Similarity 38.8%; Pred. No. 54;
Matches 19; Conservative 6; Mismatches 16; Indels 8; Gaps 3;

QY      4 TLTLOSTNTHTOSSSSSDG-GLFES-----RPAHSLP--PGEGRVERP 44
Db      49 TLTAQMNSTQSAFPAKSDAIFSSGVEGVNPIFSDEAPGYDPKDDP 97

RESULT 7
OAZ_MOUSE
ID      OAZ_MOUSE STANDARD; PRT; 226 AA.
AC      P54369; O08610;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Ornithine decarboxylase antizyme (ODC-Az).
GN      OAZ1 OR OAZ.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98088944; PubMed=9428668;
RA      Nilsson J., Koskineniemi S., Persson K., Grahn B., Holm I.;
RT      "Polyamines regulate both transcription and translation of the gene

```

```

RT      encoding ornithine decarboxylase antizyme in mouse.";
RL      Eur. J. Biochem. 250:223-231(1997).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=129/5v0;
RA      Kanare K., Uusi-Oukari M., Janne O.A.;
RL      Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: BINDS TO, AND DESTABILIZES, ORNITHINE DECARBOXYLASE
CC      WHICH IS THEN DEGRADED. ALSO INHIBITS CELLULAR UPTAKE OF
CC      POLYAMINES BY INACTIVATING THE POLYAMINE UPTAKE TRANSPORTER.
CC      -1- MISCELLANEOUS: A RIBOSOMAL FRAMESHIFT OCCURS BETWEEN THE CODONS
CC      FOR SER-67 AND ASP-68. AN AUTOREGULATORY MECHANISM ENABLES
CC      MODULATION OF FRAMESHIFTING ACCORDING TO THE CELLULAR
CC      CONCENTRATION OF POLYAMINES.
CC      -1- SIMILARITY: BELONGS TO THE ODC ANTIZYME FAMILY.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL; U52822; AAB96329.1; -.
DR      EMBL; U52823; AAB96330.1; -.
DR      EMBL; U84291; AAC53107.1; -.
DR      WGI; WGI:109433; Oa21.
DR      InterPro; IPR02993; ODC_AZ.
DR      Pfam; PF02100; ODC_AZ; 1.
DR      ProDom; PD007483; ODC_AZ; 1.
DR      PROSITE; PS01337; ODC_AZ; 1.
KM      Ribosomal frameshift.
FT      INIT MET 0
FT      CONFLICT 68 68 D -> C (IN REF. 1; AAB96330).
SQ      SEQUENCE 226 AA; 25002 MW; C580D8B730318F6 CRC64;

Query Match      16.7%; Score 60; DB 1; Length 226;
Best Local Similarity 31.7%; Pred. No. 7.3;
Matches 19; Conservative 8; Mismatches 23; Indels 10; Gaps 2;

QY      1 QSRP-----LTLOSTNTHQS-----SSSSDGLFRSRPAHSLPGEGRVERPVDAE 50
Db      104 EEPSTNDKTRVLSITSTLTKQYTRAWWSGGGLYIEPAGLPBGSKDSFALLERPAE 163

RESULT 8
ALSI_CANAL
ID      ALSI_CANAL STANDARD; PRT; 1260 AA.
AC      P46550;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Agglutinin-like protein 1 precursor.
GN      ALS1.
OS      Candida albicans (Yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX      NCBI_TaxID=5476;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 11651 / B792;
RX      MEDLINE=9572392; PubMed=7752895;
RA      Hoyer L.L., Scherer S., Shatzman A.R., Livi G.P.;
RT      "Candida albicans ALS1 domains related to a Saccharomyces cerevisiae
RT      sexual agglutinin separated by a repeating motif.".
RL      Mol. Microbiol. 15:39-54(1995).
CC      -1- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC      -1- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC      -1- SIMILARITY: TO YEAST SAG1.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L25902; AAC41649.2; Repeat; Signal.  
CELL adhesion; Glycoprotein; POTENTIAL.  
SIGNA 1 17  
FT CHAIN 18 1260  
FT DOMAIN 433 792  
FT REPEAT 433 468  
FT REPEAT 469 504  
FT REPEAT 505 540  
FT REPEAT 541 576  
FT REPEAT 577 612  
FT REPEAT 613 648  
FT REPEAT 649 684  
FT REPEAT 685 720  
FT REPEAT 721 756  
FT REPEAT 757 792  
FT DOMAIN 983 1152  
FT REPEAT 983 1043  
FT REPEAT 1092 1152  
FT DOMAIN 399 404  
FT DOMAIN 408 418  
FT DOMAIN 450 455  
FT DOMAIN 486 491  
FT DOMAIN 522 527  
FT DOMAIN 558 563  
FT DOMAIN 594 599  
FT DOMAIN 630 635  
FT DOMAIN 666 671  
FT DOMAIN 702 707  
FT DOMAIN 738 743  
FT DOMAIN 774 779  
FT DOMAIN 874 877  
FT CARBOHYD 471 471  
FT CARBOHYD 579 579  
FT CARBOHYD 615 615  
FT CARBOHYD 687 687  
FT CARBOHYD 723 723  
FT CARBOHYD 820 820  
FT CARBOHYD 886 886  
FT CARBOHYD 918 918  
FT CARBOHYD 973 973  
FT CARBOHYD 1045 1045  
FT CARBOHYD 1068 1068  
SQ SEQUENCE 1260 AA; 132641 MW; 763D1063A2354C24 CRC64;

Query Match 16.7%; Score 60; DB 1; Length 1260;  
Best Local Similarity 24.2%; Pred. No. 57;  
Matches 23; Conservative 9; Mismatches 35; Indels 28; Gaps 3;

Qy 2 SPITLT-----QSTNTHTQSSSSDGLFRSRPAHSL-----PP 36  
Db 651 NPVITTEWWSQYATTTITAPPEIDTVLIREPPNHTTTEYWSQYATTTTATPP 710

Qy 37 GEDGRV---EPYVDFAEFVRLWSVDHGQSVVTP 68  
Db 711 GETDVTLIREPPNHTTTEYWSQYATTTTATPP 745

RESULT 9  
CAHX\_FLAPR STANDARD; PRT; 329 AA.  
AC P46281;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Carbonic anhydrase 1 (BC 4.2.1.1) (Carbonate dehydratase).  
OS Flaveria pringlei.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;  
OC Heleniidae; Flaveria.  
OX NCBI\_TaxID=4226;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Leaf;  
RX MEDLINE=96046753; PubMed=7579185;  
RA Ludwig M., Burnell J.N.;  
RT "Molecular comparison of carbonic anhydrase from Flaveria species demonstrating different photosynthetic pathways.";  
RL Plant Mol. Biol. 29:353-365(1995).  
CC 1- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.  
CC 1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.  
CC 1- SUBUNIT: HOMOHXAMER (BY SIMILARITY).  
CC 1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC 1- DOMAIN: POSSESSES A TRANSIT-LIKE PEPTIDE, BUT IT IS PROPOSED THAT THIS PEPTIDE IS NOT REMOVED AND THAT THEREFORE THE ENZYME STAYS IN THE CYTOPLASM INSTEAD OF GOING TO THE CHLOROPLAST (BY SIMILARITY).  
CC 1- SIMILARITY: BELONGS TO THE PLANT AND PROKARYOTIC CARBONIC ANHYDRASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U19737; AAA86992.1; --  
DR InterPro; IPR001765; Prok\_Coanhd.  
DR Pfam; PF00484; Pro\_CA; 1.  
DR PROSITE; PS00704; PROK\_CO2\_ANHYDRASE\_1; 1.  
DR PROSITE; PS00705; PROK\_CO2\_ANHYDRASE\_2; 1.  
KW Lyase; Zinc.  
FT DOMAIN 1 108 CHLOROPLAST TRANSIT PEPTIDE-LIKE.  
FT DOMAIN 23 26 POLY-SER.  
FT DOMAIN 40 46 POLY-SER.  
SQ SEQUENCE 329 AA; 35486 MW; B18E56B1B94C34B CRC64;

Query Match 16.4%; Score 59; DB 1; Length 329;  
Best Local Similarity 27.0%; Pred. No. 15;  
Matches 20; Conservative 13; Mismatches 33; Indels 8; Gaps 2;

Qy 1 QSPITLTQSTNTHTQSSSSDGLFRSRPAHSLPP-----GEDGRVPEYVDFAEFVR 53  
Db 28 RGVLSARTCNSSSSSSATPPSLIRNEPVFAAPAIITPNWTDGN-ESVEEAIDALK 86

Qy 54 LMSVDHGQSVVTA 67  
Db 87 KMLIEKGELEPVAA 100

RESULT 10  
CAHI\_FLALI STANDARD; PRT; 330 AA.  
AC P46512;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Carbonic anhydrase 1 (BC 4.2.1.1) (Carbonate dehydratase 1).  
OS Flaveria linearis.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;  
OC Heleniidae; Flaveria.  
OX NCBI\_TaxID=4225;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Leaf;  
RX MEDLINE=96046753; PubMed=7579185;



```

CC -----
DR EMBL; AF108210; AAD19642.1; -.
DR EMBL; AF108209; AAD19639.1; -.
DR EMBL; AF108207; AAD19639.1; JOINED.
DR EMBL; AF108208; AAD19639.1; JOINED.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCRHOOPS.
DR PROSITE; PS00237; G-PROTEIN RECP FL 1; 1.
DR PROSITE; PS00262; G-PROTEIN RECP FL 2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Lipoprotein; Palmitate.
FT DOMAIN 1 41 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 42 63 1 (POTENTIAL).
FT DOMAIN 64 82 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 83 103 2 (POTENTIAL).
FT DOMAIN 104 121 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 122 143 3 (POTENTIAL).
FT DOMAIN 144 163 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 164 184 4 (POTENTIAL).
FT DOMAIN 185 220 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 221 241 5 (POTENTIAL).
FT DOMAIN 242 272 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 273 293 6 (POTENTIAL).
FT DOMAIN 294 313 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 314 333 7 (POTENTIAL).
FT DOMAIN 334 399 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 10 10 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 29 29 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 120 203 BY SIMILARITY.
SQ SEQUENCE 399 AA; 44373 MW; B48DD27197AED02EB CRC64;

Query Match 16.4%; Score 59; DB 1; Length 399;
Best Local Similarity 34.7%; Pred. No. 19;
Matches 17; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

QY 1 OSPTLTLOSTWTHQSSSSDGLFRSPAHSLPRGDRVPEYVDF 49
   |||||
DB 7 QSFNOLISTNDIESSSVVPNDSTNKRTGDNFGLCALCIYIYA 55

RESULT 13
PRLR MOUSE STANDARD; PRT; 608 AA.
AC Q08501; Q62099; P15213; P15212;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Prolactin receptor precursor (PRL-R).
GN PRLR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (FORM PRL-R3).
RC STRAIN=C3H; TISSUE=Mammary gland;
RX MEDLINE=94085788; PubMed=8262385;
RA Moore R.C., Oka T.;
RT "Cloning and sequencing of the cDNA encoding the murine mammary gland
RT long-form prolactin receptor.";
RL Gene 134:263-265(1993).
RN [2]
RP SEQUENCE FROM N.A. (FORM PRL-R3).
RC STRAIN=Swiss Webster; TISSUE=Liver;
RX MEDLINE=93307149; PubMed=8319571;
RA Clarke D.L., Linzer D.I.H.;
RT "Changes in prolactin receptor expression during pregnancy in the
RT mouse ovary.";
RL Endocrinology 133:224-232(1993).
RN [3]
RP SEQUENCE FROM N.A. (FORM PRL-R3).

```

```

RA Sasaki M.;
RL Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A. (FORM PRL-R3).
RC STRAIN=BALB/c; TISSUE=Mammary gland;
RA Edery M., Pezet A., Nandi S., Kelly P.A.;
RN Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A. (FORMS PRL-R2 AND PRL-R1).
RC STRAIN=Swiss Webster; TISSUE=Liver;
RX MEDLINE=89261824; PubMed=2725531;
RA Davis J.A., Linzer D.I.H.;
RT "Expression of multiple forms of the prolactin receptor in mouse
RT liver.";
RL Mol. Endocrinol. 3:674-680(1989).
CC -!- FUNCTION: THIS IS A RECEPTOR FOR THE ANTERIOR PITUITARY HORMONE
CC PROLACTIN, AS WELL AS PLACENTAL LACTOGEN I AND II.
CC -!- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; PRL-R1, PRL-R2 AND PRL-R3 (SHOWN
CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L13593; AAC37641.1; -.
DR EMBL; L14811; AAA02686.1; -.
DR EMBL; D10214; BAA01066.1; -.
DR EMBL; X73372; CAA51789.1; -.
DR EMBL; M22959; AAA39977.1; -.
DR EMBL; M22958; AAA39976.1; -.
DR PIR; JT0671; JT0671.
DR HSP; P16471; 1BP3.
DR MGD; MGI:97763; Prlr.
DR InterPro; IPR002996; CRJA.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003528; Hemtopoptn_L_F1.
DR Pfam; PF00041; fn3; 2.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS01352; HEMATOPO REC L_F1; 1.
KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat;
KW Alternative splicing.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 608 PROLACTIN RECEPTOR.
FT DOMAIN 20 229 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 230 253 BY SIMILARITY.
FT DOMAIN 254 608 CYTOPLASMIC (BY SIMILARITY).
FT DOMAIN 20 117 FIBRONECTIN TYPE-III 1.
FT DOMAIN 119 222 FIBRONECTIN TYPE-III 2.
FT DISULFID 31 41 BY SIMILARITY.
FT DISULFID 70 81 BY SIMILARITY.
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 281 292 KKGSEELLSALG -> VHNKEQENTVY (IN ISOFORM
FT PRL-R2).
FT VARSPPLIC 293 608 MISSING (IN ISOFORM PRL-R2).
FT VARSPPLIC 281 303 KKGSEELLSALGQDFPTSDCE -> LMCSTLQUTSLVKI
FT PTFLECLDL (IN ISOFORM PRL-R1).
FT VARSPPLIC 304 608 MISSING (IN ISOFORM PRL-R1).
FT CONFLICT 558 558 L -> F (IN REF. 2).
SQ SEQUENCE 608 AA; 68240 MW; B8CE202B2EFC9FC6 CRC64;

Query Match 16.3%; Score 58.5; DB 1; Length 608;
Best Local Similarity 24.7%; Pred. No. 35;
Matches 18; Conservative 12; Mismatches 24; Indels 19; Gaps 2;

```



```

QY 2 SPTLLSTNTHTQSSSSDGLFRSPAHSLPAGEDGVPEYVDFAPFRLMS----- 56
D 388 TPNNQNTPNCHNDTSTKST-----WPLPQCHNRSPHSIADVCKLKGSPEDT 437
D 57 ----VDHGEQSV 65
D 438 LDSFLDKAEENVL 450

RESULT 14
AF17_HUMAN STANDARD; PRT; 1093 AA.
AC P55198;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE AF-17 protein.
GN MULT6 OR AF17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94336695; PubMed=8059765;
RA Prasad R., Leshkowitz D., Gu Y., Alder H., Nakamura T., Saito H.,
RA Huebner K., Berger R., Croce C.M., Canaan E.;
RA "Leucine-zipper dimerization motif encoded by the AF17 gene fused to
RT AL1-1 (ML) in acute leukemia."
RT Proc. Natl. Acad. Sci. U.S.A. 91:8107-8111(1994).
RL CELLULAR LOCATION: Nuclear (Potential).
CC -1- DISSEMINATED INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL
CC TRANSLOCATION T(11;17)(Q23;Q21) THAT INVOLVES MULT6 AND MLL/HRX.
CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.
CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC -1- SIMILARITY: HIGH, TO AF10.
CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/AF17.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U07932; AAA21145.1; -.
DR Genbank; HGNC:7138; MLLT6.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 2.
KW Nuclear protein; Zinc-finger; Repeat; Chromosomal translocation;
KW Proto-oncogene.
FT ZN_FING 5 57 PHD-TYPE 1.
FT ZN_FING 51 68 C4-TYPE.
FT ZN_FING 117 180 PHD-TYPE 2.
FT DOMAIN 190 211 GLY/SER-RICH.
FT DOMAIN 275 282 POLY-SER.
FT DOMAIN 326 338 POLY-SER.
FT DOMAIN 729 764 LEUCINE-ZIPPER.
FT DOMAIN 822 829 PRO-SER.
FT DOMAIN 834 862 PRO-RICH.
FT DOMAIN 935 984 GLN-RICH.
FT DOMAIN 1040 1051 POLY-ALA.
FT DOMAIN 1069 1080 GLY-RICH.
FT SITE 551 MLL FUSION POINT (IN ACUTE MYELOID
FT LEUKEMIA PATIENT).
SQ SEQUENCE 1093 AA; 112021 MW; F60042A6D3BF579E CRC64;

```

```

QY 14 TQSSSSDGLFRSPAHSLPAGEDGVPEYVDFAPFRLMSVHGEQSVTAP 68
D 330 SSSSSSSSGGPF--QAVS-----SLQSPDFSAFPLKLPQEPEDKYSKPTAP 375

Query Match 16.3%; Score 58.5; DB 1; Length 1093;
Best Local Similarity 36.4%; Pred. No. 71;
Matches 20; Conservative 8; Mismatches 18; Indels 9; Gaps 2;

RESULT 15
PLC1_CANAL STANDARD; PRT; 1099 AA.
ID PLC1_CANAL
AC O13433;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase 1
DE (EC 3.1.4.11) (PLC-1) (Phospholipase C-1).
GN PLC1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=132A;
RA MEDLINE=96129081; PubMed=9467900;
RA Bennett D.E., McCreary C.E., Coleman D.C.;
RA "Genetic characterization of a phospholipase C gene from Candida
RT albicans: presence of homologous sequences in Candida species other
RT than Candida albicans."
RT Microbiology 144:55-72(1998).
CC -1- FUNCTION: THE PRODUCTION OF THE SECOND MESSENGER MOLECULES
CC DIACYLGLYCEROL (DAG) AND INOSITOL 1,4,5-TRISPHOSPHATE (IP3) IS
CC MEDIATED BY ACTIVATED PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE
CC C ENZYMES.
CC -1- CATALYTIC ACTIVITY: 1-phosphatidyl-1D-myo-inositol 4,5-
CC bisphosphate + H(2)O = D-myo-inositol 1,4,5-trisphosphate +
CC diacylglycerol.
CC -1- SIMILARITY: DOMAINS X AND Y ARE CONSERVED IN DIFFERENT FORMS OF
CC PLC AND ARE ESSENTIAL FOR CATALYTIC ACTIVITY.
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y13975; CAA74308.1; -.
DR HSRP; P10688; IDIX.
DR InterPro; IPR000009; C2.
DR InterPro; IPR001192; PI_PLC.
DR InterPro; IPR000909; PI_PLC_Xdom.
DR InterPro; IPR001711; PI_PLC_Y.
DR Pfam; PF00387; PI_PLC_X; 1.
DR Pfam; PF00388; PI_PLC_Y; 1.
DR PRINTS; PR00390; PHPLIPASEC.
DR ProDom; PD001202; PI_PLC_Y; 1.
DR SMART; SM00239; C2; 1.
DR SMART; SM00148; PLCX; 1.
DR SMART; SM00149; PLCY; 1.
DR PROSITE; PS50004; C2 DOMAIN 2; FALSE_NEG.
DR PROSITE; PS50007; PIPLC_X DOMAIN; 1.
DR PROSITE; PS50008; PIPLC_Y DOMAIN; 1.
KW Hydrolase; Lipid degradation; Transducer.
FT DOMAIN 566 726 DOMAIN X.
FT DOMAIN 794 912 DOMAIN Y.
FT DOMAIN 938 1066 C2 DOMAIN.
FT ACT_SITE 579 BY SIMILARITY.
FT ACT_SITE 642 BY SIMILARITY.

```

```
FT DOMAIN          90 102 POLY-SER.
FT DOMAIN        239 243 POLY-THR.
FT DOMAIN        517 520 POLY-ASP.
FT DOMAIN        619 623 POLY-THR.
FT DOMAIN        629 632 POLY-ASP.
FT DOMAIN        743 760 POLY-THR.
FT DOMAIN        918 921 POLY-SER.
SQ SEQUENCE 1099 AA; 124591 MW; D54D687D53A2829B CRC64;

Query Match      16.3%; Score 58.5; DB 1; Length 1099;
Best Local Similarity 28.1%; Pred.No. 72;
Matches 16; Conservative 12; Mismatches 20; Indels 9; Gaps 3;

QY 1 QSPRTLQSTNTHQTSSSSSDGGLFRSR---PAHSIP--PGEDGR--VEPYVDF 48
Db 912 RKPTLKSSSNVDTRISLTINSKTIRNFELISGHQLPRFPDDYKDQAINPYSF 968
```

Search completed: December 10, 2002, 07:31:39  
Job time : 45 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 07:26:16 (Search time 96 Seconds)

145.950 Million cell updates/sec

Title: US-09-830-144-4\_COPY\_437\_504

Perfect score: 359 1 QSPFTLLQSTNTHQTSSSSSS.....AEFYRLWSVDHGQSVVTAP 68

Sequence: 1 QSPFTLLQSTNTHQTSSSSSS.....AEFYRLWSVDHGQSVVTAP 68

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP archaea.\*  
2: SP bacteria.\*  
3: SP fungi.\*  
4: SP human.\*  
5: SP invertebrate.\*  
6: SP mammal.\*  
7: SP mhc.\*  
8: SP organelle.\*  
9: SP phage.\*  
10: SP plant.\*  
11: SP rodent.\*  
12: SP virus.\*  
13: SP vertebrate.\*  
14: SP unclassified.\*  
15: SP rivirus.\*  
16: SP bacteriap.\*  
17: SP archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	356	99.2	500	11 Q8R0D1	Q8R0D1 mus musculu
2	275	76.6	52	11 Q9CV62	Q9CV62 mus musculu
3	260.5	72.6	498	13 Q73614	Q73614 xenopus lae
4	74.5	20.8	329	4 Q96S04	Q96S04 homo sapien
5	70	19.5	1162	5 Q9VMS7	Q9VMS7 dirosophila
6	69.5	19.4	389	12 Q9DXA1	Q9DXA1 avian pneum
7	69.5	19.4	389	12 Q9DXA0	Q9DXA0 avian pneum
8	67.5	18.8	703	4 Q9H4S1	Q9H4S1 homo sapien
9	67.5	18.8	739	4 Q43584	Q43584 homo sapien
10	67.5	18.8	739	4 Q9BY75	Q9BY75 homo sapien
11	67.5	18.8	903	4 Q96F66	Q96F66 homo sapien
12	66	18.4	591	16 Q9RCD3	Q9RCD3 streptomyce
13	65.5	18.2	634	11 Q92508	Q92508 mus musculu
14	65	18.1	452	12 Q91TC8	Q91TC8 hepatitis c
15	65	18.1	452	12 Q91TC1	Q91TC1 hepatitis c
16	65	18.1	452	12 Q91TB9	Q91TB9 hepatitis c

17	64.5	18.0	673	3 Q9C2H6	Q9C2H6 neurospora
18	64	17.8	452	12 Q91TC6	Q91TC6 hepatitis c
19	64	17.8	5604	4 Q8W253	Q8W253 homo sapien
20	63.5	17.7	780	13 Q9PU14	Q9PU14 xenopus lae
21	63.5	17.7	3021	12 Q92933	Q92933 hepatitis c
22	63	17.5	452	12 Q91TD5	Q91TD5 hepatitis c
23	63	17.5	452	12 Q91TD3	Q91TD3 hepatitis c
24	63	17.5	452	12 Q91TC4	Q91TC4 hepatitis c
25	63	17.5	452	12 Q91TC3	Q91TC3 hepatitis c
26	63	17.5	452	12 Q91TB8	Q91TB8 hepatitis c
27	63	17.5	452	12 Q91TB6	Q91TB6 hepatitis c
28	63	17.5	452	12 Q91TB4	Q91TB4 hepatitis c
29	63	17.5	2112	5 Q9VEL9	Q9VEL9 dirosophila
30	63	17.5	3021	12 Q81258	Q81258 hepatitis c
31	63	17.5	3021	12 Q68870	Q68870 hepatitis c
32	62.5	17.4	358	4 Q96E87	Q96E87 homo sapien
33	62.5	17.4	1656	10 Q91G61	Q91G61 oryza sativ
34	62.5	17.4	1945	5 Q9V768	Q9V768 dirosophila
35	62	17.3	172	10 Q9P6W8	Q9P6W8 arabidopsis
36	62	17.3	639	4 Q9EP30	Q9EP30 homo sapien
37	62	17.3	679	16 Q97SL8	Q97SL8 streptococ
38	62	17.3	734	4 Q96EP31	Q96EP31 homo sapien
39	62	17.3	734	4 Q961A4	Q961A4 homo sapien
40	62	17.3	740	4 Q96EP29	Q96EP29 homo sapien
41	62	17.3	3583	5 Q9V3N4	Q9V3N4 dirosophila
42	62	17.3	4957	4 Q14687	Q14687 homo sapien
43	62	17.3	5262	4 Q14686	Q14686 homo sapien
44	61.5	17.1	188	3 Q96U43	Q96U43 neurospora
45	61.5	17.1	532	5 Q9BKYS	Q9BKYS leishmania

## ALIGNMENTS

RESULT 1  
ID Q8R0D1 PRELIMINARY; PRT; 500 AA.

AC Q8R0D1;  
DT 01-UN-2002 (TREMUREL. 21, Last sequence update)  
DT 01-UN-2002 (TREMUREL. 21, Last annotation update)  
DE Similar to mitogen-activated protein kinase kinase 7  
DE interacting protein 1 (Fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC027054; AAH27054.1; -.  
KW kinase.  
FT NON-TER.  
SQ SEQUENCE 500 AA; 54413 MW; 97529D2E2AE4DA34 CRC64;

Query Match 99.2%; Score 356; DB 11; Length 500;  
Best Local Similarity 98.5%; Pred. No. 5.7e-35;  
Matches 67; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QSPFTLLQSTNTHQTSSSSDGLFRSPRAHSIPREDGRVPEPYDFAEFRLMSVDHG 60  
Db 433 QSPFTLLQSTNTHQTSSSSDGLFRSPRAHSIPREDGRVPEPYDFAEFRLMSVDHG 492

Qy 61 EQSVVTAP 68  
Db 493 EQSVVTAP 500

RESULT 2  
Q9CV62 PRELIMINARY; PRT; 52 AA.

```

AC Q9CV62;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 2310012M03Rik protein (Fragment).
GN 2310012M03Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa K., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bulc C., Fleischer C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyoko-oka K., Wang K.H., Weitz C., Whittaker C., Wilmink L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK009321; BAB26215.1;
DR MGD; MGI:1913763; 2310012M03Rik.
FT NON TER 1
SQ SEQUENCE 52 AA; 5712 MW; 813E29B1639920A6 CRC64;

Query Match 76.6%; Score 275; DB 11; Length 52;
Best Local Similarity 98.1%; Pred. No. 2.9e-26;
Matches 51; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 17 SSSSSDGGFLFRSPAHSLPPGSDGRVEPYVDFAFBYRLWSVDHGQSVWTPAP 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 SSSSSDGGFLFRSPAHSLPPGSDGRVEPYVDFAFBYRLWSVDHGQSVWTPAP 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 3
O73614
ID O73614 PRELIMINARY; PRT; 498 AA.
AC O73614;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE TAB1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98130593; PubMed=9463380;
RA Shibuya H., Iwata H., Masuyama N., Gotoh Y., Yamaguchi K., Irie K.,
RA Matsumoto K., Nishida E., Ueno N.;
RT "Role of TAK1 and TAB1 in BMP signaling in early Xenopus
development.";
RL EMBO J. 17:1019-1028(1998).
DR EMBL; U920311; AAC14009.1;
DR InterPro; IPR001932; PP2C-like.
DR Pfam; PF00481; PP2C; 1.
DR SMART; SM00332; PP2CC; 1.

```

```

SQ SEQUENCE 498 AA; 54450 MW; 654AB8DE2353524 CRC64;

Query Match 72.6%; Score 260.5; DB 13; Length 498;
Best Local Similarity 73.5%; Pred. No. 2.4e-23;
Matches 50; Conservative 8; Mismatches 9; Indels 1; Gaps 1;

QY 1 QSTLTQLQSTNTHTQSSSSSDGGFLFRSPAHSLPPGSDGRVEPYVDFAFBYRLWSVDHG 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 430 QSPSATLQSTNTHTQSSSSSDGGFLFRSPAHSLPPGSDGRVEPYVDFAFBYRLWSVDHG 499
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 EQ-SVVTA 67
DB 490 DPGTLLTA 497

RESULT 4
O96S04
ID O96S04 PRELIMINARY; PRT; 329 AA.
AC O96S04;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 35.8 kDa protein.
GN GS78.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21096910; PubMed=1157797;
RA Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA Higgs D.R.;
RT "Sequence, structure and pathology of the fully annotated terminal 2
KT Mb of the short arm of human chromosome 16.";
RL Hum. Mol. Genet. 10:339-352(2001).
DR EMBL; AE006465; AAK61262.1;
KW Hypothetical protein.
SQ SEQUENCE 329 AA; 35799 MW; 890FE4B3D1C5976D CRC64;

Query Match 20.8%; Score 74.5; DB 4; Length 329;
Best Local Similarity 28.6%; Pred. No. 0.65;
Matches 22; Conservative 8; Mismatches 28; Indels 19; Gaps 2;

QY 2 SPILTQLQSTNTHTQSSSSSDGGFLFRSPAHSLPPGSDG-----RVPEY 45
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 138 APTTRICRTRCPQSGSLSDG---RNRCPDASESNHGRPHGSSPVLYGFIRICRVERN 194
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 46 VDFAFYRLWSVDHGQ 62
DB 195 IPECEDFTWTLGSGEK 211

RESULT 5
O9VWB7
ID O9VWB7 PRELIMINARY; PRT; 1162 AA.
AC O9VWB7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG11940 protein.
GN CG11940.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceilniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

```

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazel R.G., Champe M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abtali J.F., Abghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Bau A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck P., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dudin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira G.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusaker D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitek R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RT Science 287:2185-2195 (2000).  
 DR EMBL: AF003513; AAF4029.1; -  
 DR F1Base; FBN0031079; CG11940.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000159; RA\_domain.  
 DR Pfam; PF00169; PH; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00314; RA; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR PRINTS; PRO1217; PRICHEXTENSN.  
 SQ SEQUENCE 1162 AA; 124386 MW; CA4C25D1E9AD795 CRC64;

Query Match 19.5%; Score 70; DB 5; Length 1162;  
 Best Local Similarity 32.8%; Pred. No. 10;  
 Matches 22; Conservative 9; Mismatches 24; Indels 12; Gaps 2;

QY 3 PTLTLOSTNTHTTSSSSSDGGLFRS-RPAHSLPPGDEGVVEVPFAERYRUMSDHCE 61  
 Db 960 PVLFPORPSTLTSCHSSSSAGSAVQYVAPGPMPLPFR-----ADVARLUSLSNGS 1008  
 QY 62 QSVYVAP 68  
 Db 1009 SSEVTSP 1015

RESULT 6  
 Q9DXA1 PRELIMINARY; PRT; 389 AA.  
 AC Q9DXA1;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Attachment membrane glycoprotein.  
 GN G.  
 OS Avian pneumovirus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Pneumovirinae; Metapneumovirus.  
 NCBI\_TaxID=38525;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FR/85/1;  
 RX MEDLINE=20495153; PubMed=11038385;  
 RA Bayon-Abboyer M.H., Arnould C., Toguin D., Eterradossi N.;  
 RT "Nucleotide sequences of the F, L and G protein genes of two non-  
 RL A/non-B avian pneumoviruses (APV) reveal a novel APV subgroup.";  
 DR J. Gen. Virol. 81:2723-2733 (2000).  
 DR EMBL: AJ251085; CAC13041.1; -  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR PRINTS; PRO1217; PRICHEXTENSN.  
 SQ SEQUENCE 389 AA; 41838 MW; 755CB4DDEC9C3EB8 CRC64;

Query Match 19.4%; Score 69.5; DB 12; Length 389;  
 Best Local Similarity 37.2%; Pred. No. 3.2;  
 Matches 16; Conservative 7; Mismatches 17; Indels 3; Gaps 1;

QY 1 QSPPTLTOSTNTHTTSSSSSDGGLFRS-RPAHSLPPGDEGRVE 43  
 Db 274 RSPPTKROETGRATPRNTATTOG---SSPPHSSPPGVADNME 313

RESULT 7  
 Q9DXA0 PRELIMINARY; PRT; 389 AA.  
 AC Q9DXA0;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Attachment membrane glycoprotein.  
 GN G.  
 OS Avian pneumovirus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Pneumovirinae; Metapneumovirus.  
 NCBI\_TaxID=38525;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FR/85/2;  
 RX MEDLINE=20495153; PubMed=11038385;  
 RA Bayon-Abboyer M.H., Arnould C., Toguin D., Eterradossi N.;  
 RT "Nucleotide sequences of the F, L and G protein genes of two non-  
 RL A/non-B avian pneumoviruses (APV) reveal a novel APV subgroup.";  
 DR J. Gen. Virol. 81:2723-2733 (2000).  
 DR EMBL: AJ288946; CAC13042.1; -  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR PRINTS; PRO1217; PRICHEXTENSN.  
 SQ SEQUENCE 389 AA; 41812 MW; 7A54EFBBS93E9F7F CRC64;

Query Match 19.4%; Score 69.5; DB 12; Length 389;  
 Best Local Similarity 37.2%; Pred. No. 3.2;  
 Matches 16; Conservative 7; Mismatches 17; Indels 3; Gaps 1;

QY 1 QSPPTLTOSTNTHTTSSSSSDGGLFRS-RPAHSLPPGDEGRVE 43  
 Db 274 RSPPTKROETGRATPRNTATTOG---SSPPHSSPPGVADNME 313

RESULT 8  
 Q9H451 PRELIMINARY; PRT; 703 AA.  
 AC Q9H451;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE DJ46801.1 (Atrophin 1 interacting protein 4 (AIP4)) (Fragment).  
 GN DJ46801.1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith M.;



```

DR Pfam; PF00632; HECT; 1.
DR Pfam; PF00397; WW; 4.
DR PRINTS; PR00403; MWDOMAIN.
DR SMART; SM00239; C2; 1.
DR SMART; SM00119; HECT; 1.
DR SMART; SM00456; WW; 4.
DR PROSITE; PS50004; C2_DOMAIN_2; 1.
DR PROSITE; PS50237; HECT; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; 4.
DR PROSITE; PS50020; WW_DOMAIN_2; 2.
KW Ligase.
SQ SEQUENCE 862 AA; 98675 MW; A3D960E7F4DBF9D3 CRC64;

Query Match 18.8%; Score 67.5; DB 4; Length 862;
Best Local Similarity 31.1%; Pred. No. 14;
Matches 23; Conservative 12; Mismatches 20; Indels 19; Gaps 3;

QY 2 SPTLTQSTNTHTQSSSSSDG-----GLFRSRPAH-----SLPPGEDGRVEPYVDF 48
DB 242 SSTGSLPPTNTNTNTSSEGTGLIPLRTISGSGRPPLNPVTQAPLPFGWEQRVD----- 296
QY 49 AEFYRLMSVDHGEQ 62
DB 297 -QHGRVYVDHVEK 309

RESULT 11
Q96F66 PRELIMINARY; PRT; 903 AA.
ID Q96F66;
AC Q96F66;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Similar to Itchy (mouse homolog) E3 ubiquitin protein
DE ligase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011571; AAH11571.1; -.
DR InterPro; IPR000008; C2.
DR InterPro; IPR000569; HECT_domain.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00632; HECT; 1.
DR Pfam; PF00397; WW; 4.
DR PROSITE; PS50004; C2_DOMAIN_2; 1.
DR PROSITE; PS50237; HECT; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; UNKNOWN_4.
DR PROSITE; PS50020; WW_DOMAIN_2; 2.
SQ SEQUENCE 903 AA; 102802 MW; 6777A2043C7B67BC CRC64;

Query Match 18.8%; Score 67.5; DB 4; Length 903;
Best Local Similarity 31.1%; Pred. No. 15;
Matches 23; Conservative 12; Mismatches 20; Indels 19; Gaps 3;

QY 2 SPTLTQSTNTHTQSSSSSDG-----GLFRSRPAH-----SLPPGEDGRVEPYVDF 48
DB 283 SSTGSLPPTNTNTNTSSEGTGLIPLRTISGSGRPPLNPVTQAPLPFGWEQRVD----- 337
QY 49 AEFYRLMSVDHGEQ 62
DB 338 -QHGRVYVDHVEK 350

RESULT 12
Q9RKD3 PRELIMINARY; PRT; 591 AA.
ID Q9RKD3

```

```

AC Q9RKD3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative integral membrane protein.
GN SC03154 OR SCE87.05.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleeser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleeser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleeser T., Latke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL132674; CAB59650.1; -.
DR InterPro; IPR003342; PMT.
DR Pfam; PF02366; PMT; 1.
SQ SEQUENCE 591 AA; 66289 MW; DE7F1707C0071897 CRC64;

Query Match 18.4%; Score 66; DB 16; Length 591;
Best Local Similarity 47.4%; Pred. No. 14;
Matches 18; Conservative 2; Mismatches 14; Indels 4; Gaps 1;

QY 27 RSRPASHLPPEDGRVEPYVDFAEFYRL----WSVDHG 60
DB 255 RSRLAALPVEDDGRVPRDAHVAETRLGWRPRLAAG 292

RESULT 13
Q92508 PRELIMINARY; PRT; 634 AA.
ID Q92508
AC Q92508;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Dachshund-like protein DACH2.
GN DACH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

RC	STRAIN=SWISS-WEBSTER/NIH;
RA	MEDLINE=21184114; PubMed=11287190;
RX	Davis R.J., Shen W., Sandler Y.I., Heanue T.A., Mardon G.;
RT	"Characterization of mouse Dach2, a homologue of Drosophila
RT	dachshund."
RL	Mech. Dev. 102:169-179(2001).
DR	EMBL; AF257217; AAK39983.1; -.
DR	MGD; MGI:1890446; Dach2.
DR	InterPro; IPR003380; Transform_Skl.
DR	Pfam; PF02437; Ski_Sno; 1.
SQ	SEQUENCE 634 AA; 68598 MW; 838A491FF20C410 CRC64;
Query Match 18.2%; Score 65.5; DB 11; Length 634;	
Best Local Similarity 51.7%; Pred. No. 17;	
Matches 15; Conservative 4; Mismatches 9; Indels 1; Gaps 1;	
Qy	16 SSSSSDGGFLFRSRPAHSLPPGCDGRVEP 44
Db	13 SSSAGVGGFLFRAEPLYS-SGGEPRRLTP 40
RESULT 14	
Q91TC8	PRELIMINARY; PRT; 452 AA.
ID	Q91TC8
AC	Q91TC8;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE	Polyprotein (Fragment).
OS	Hepatitis C virus.
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC	Hepacivirus.
OX	NCBI_TaxID=11103;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=SR12;
RA	Castelain S., Duverlie G., Baron A.;
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF320798; AAK49495.1; -.
DR	InterPro; IPR002868; HCV_NS5a.
DR	Pfam; PF01506; HCV_NS5a; 1.
FT	NON TER 1
FT	NON TER 452 452
SQ	SEQUENCE 452 AA; 49053 MW; 49CB2949009C5985 CRC64;
Query Match 18.1%; Score 65; DB 12; Length 452;	
Best Local Similarity 38.5%; Pred. No. 13;	
Matches 25; Conservative 6; Mismatches 20; Indels 14; Gaps 4;	
Qy	9 STNTHTSSSSS----SDGGLFRSRPAHSLPP--GEDGRVEPYVDFAEFYRLWSV--DHG 60
Db	392 SSGVDTSQTSTSKVPLSPGGSDESSECSMPLEGEPPDPLSCD-----SWTIVSDNE 445
Qy	61 EQSVV 65
Db	446 EQSVV 450
RESULT 15	
Q91TC1	PRELIMINARY; PRT; 452 AA.
ID	Q91TC1
AC	Q91TC1;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE	Polyprotein (Fragment).
OS	Hepatitis C virus.
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC	Hepacivirus.
OX	NCBI_TaxID=11103;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NR2;



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 07:29:16 ; Search time 38 seconds  
(without alignments)  
52.652 Million cell updates/sec

Title: US-09-830-144-4\_COPY\_437\_504

Perfect score: 359  
Sequence: 1 GSPITLQSTVHTTQSSSS.....AEPTLMSVDHGQSVVTAP 68

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	359	100.0	504	2	US-08-752-891-2
2	359	100.0	504	2	US-08-752-891-6
3	359	100.0	504	2	US-09-144-178-2
4	359	100.0	504	2	US-09-144-178-6
5	359	100.0	504	4	US-09-406-854-2
6	359	100.0	504	4	US-09-406-854-6
7	359	100.0	504	4	US-09-529-279-2
8	359	100.0	513	4	US-09-529-279-43
9	359	100.0	517	4	US-09-529-279-11
10	77	21.4	16	4	US-09-529-279-41
11	67.5	18.8	852	2	US-09-070-060-3
12	67.5	18.8	852	3	US-09-357-746-3
13	66	18.4	16	4	US-09-529-279-40
14	63.5	17.7	240	4	US-09-152-060-100
15	63.5	17.7	335	4	US-09-152-060-64
16	63.5	17.7	391	4	US-08-706-216-6
17	60.5	16.9	854	2	US-09-070-060-4
18	60.5	16.9	854	3	US-09-357-746-4
19	60	16.7	1037	4	US-09-428-711A-21
20	58.5	16.3	1093	5	US-08-545-860B-55
21	58.5	16.3	1093	5	PCT-US94-04496-55
22	57.5	16.0	421	4	US-09-322-478-6
23	57.5	16.0	856	4	US-09-152-060-77
24	56.5	15.7	922	4	US-09-116-473-4
25	56.5	15.7	1385	2	US-08-687-399-7
26	56	15.6	181	1	US-08-278-091-15
27	56	15.6	181	1	US-08-483-859-15

28	56	15.6	181	1	US-08-472-173-15	Sequence 15, Appl
29	56	15.6	181	2	US-08-487-167-15	Sequence 15, Appl
30	56	15.6	181	2	US-08-482-816-15	Sequence 15, Appl
31	56	15.6	181	2	US-08-296-148-15	Sequence 15, Appl
32	56	15.6	181	2	US-08-801-499-15	Sequence 15, Appl
33	56	15.6	181	2	US-08-615-271-15	Sequence 15, Appl
34	56	15.6	181	3	US-09-074-660-15	Sequence 15, Appl
35	56	15.6	181	3	US-09-074-659-15	Sequence 15, Appl
36	56	15.6	181	3	US-09-106-468-15	Sequence 15, Appl
37	56	15.6	181	4	US-09-106-468A-15	Sequence 15, Appl
38	56	15.6	181	4	US-09-106-467-15	Sequence 15, Appl
39	56	15.6	297	6	5514590-2	Patent No. 5514590
40	56	15.6	305	4	US-08-904-234-1	Sequence 1, Appl
41	56	15.6	315	3	US-08-965-903B-8	Sequence 8, Appl
42	56	15.6	315	4	US-09-370-396-3	Sequence 3, Appl
43	56	15.6	427	4	US-09-086-483A-4	Sequence 4, Appl
44	56	15.6	427	4	US-09-041-886-2	Sequence 2, Appl
45	56	15.6	427	4	US-09-006-353A-5	Sequence 5, Appl

## ALIGNMENTS

RESULT 1  
US-08-752-891-2  
Sequence 2, Application US/08752891  
Patent No. 5837819  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunhiro  
APPLICANT: NISHIDA, Eisuke  
TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,891  
FILING DATE: 20-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TRIEUX: 904136  
INFORMATION FOR SEQ ID NO. 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 504 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-752-891-2  
Query Match 100.0%; Score 359; DB 2; Length 504;  
Best Local Similarity 100.0%; Pred. No. 2.3e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QSPFTLTQSTNTHTQSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDFAEFYRLWSVDHG 60  
Db 437 QSPFTLTQSTNTHTQSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDFAEFYRLWSVDHG 496

Qy 61 EQSVVTAP 68  
Db 497 EQSVVTAP 504

## RESULT 2

US-08-752-891-6  
; Sequence 6, Application US/08752891  
; Patent No. 5837819  
; GENERAL INFORMATION:  
; APPLICANT: MATSUMOTO, Kunihiro  
; APPLICANT: NISHIDA, Eisuke  
; TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08752,891  
; FILING DATE: 20-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-300856  
; FILING DATE: 28-OCT-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-126282  
; FILING DATE: 24-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 17981/111  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 504 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-752-891-6

Query Match 100.0%; Score 359; DB 2; Length 504;  
Best Local Similarity 100.0%; Pred. No. 2.3e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QSPFTLTQSTNTHTQSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDFAEFYRLWSVDHG 60  
Db 437 QSPFTLTQSTNTHTQSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDFAEFYRLWSVDHG 496

Qy 61 EQSVVTAP 68  
Db 497 EQSVVTAP 504

## RESULT 3

US-09-144-178-2  
; Sequence 2, Application US/09144178

; Patent No. 5989862  
; GENERAL INFORMATION:  
; APPLICANT: MATSUMOTO, Kunihiro  
; APPLICANT: NISHIDA, Eisuke  
; TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09144,178  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/752,891  
; FILING DATE: 20-NOV-1996  
; APPLICATION NUMBER: JP 8-300856  
; FILING DATE: 28-OCT-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-126282  
; FILING DATE: 24-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 17981/111  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 504 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-144-178-2

Query Match 100.0%; Score 359; DB 2; Length 504;  
Best Local Similarity 100.0%; Pred. No. 2.3e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QSPFTLTQSTNTHTQSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDFAEFYRLWSVDHG 60  
Db 437 QSPFTLTQSTNTHTQSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDFAEFYRLWSVDHG 496

Qy 61 EQSVVTAP 68  
Db 497 EQSVVTAP 504

## RESULT 4

US-09-144-178-6  
; Sequence 6, Application US/09144178  
; Patent No. 5989862  
; GENERAL INFORMATION:  
; APPLICANT: MATSUMOTO, Kunihiro  
; APPLICANT: NISHIDA, Eisuke  
; TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.

COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/144,178  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/752,891  
FILING DATE: 20-NOV-1996  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 504 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-144-178-6

Query Match 100.0%; Score 359; DB 2; Length 504;  
Best Local Similarity 100.0%; Pred. No. 2.3e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 OSPTLTQSTNHTQSSSSSDGGLFRSPAHSLPPGEDGRVEPYVDFAEFRYRLMSYDHG 60  
DB 437 OSPTLTQSTNHTQSSSSSDGGLFRSPAHSLPPGEDGRVEPYVDFAEFRYRLMSYDHG 496

QY 61 EOSVVTAP 68  
DB 497 EOSVVTAP 504

RESULT 5  
US-09-406-854-2  
Sequence 2, Application US/09406854  
Patent No. 6140042  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihiko  
APPLICANT: NISHIDA, Eisuke  
TITLE OF INVENTION: TABI PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/406,854  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,891  
FILING DATE: 20-NOV-1996  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 504 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-406-854-2

Query Match 100.0%; Score 359; DB 4; Length 504;  
Best Local Similarity 100.0%; Pred. No. 2.3e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 OSPTLTQSTNHTQSSSSSDGGLFRSPAHSLPPGEDGRVEPYVDFAEFRYRLMSYDHG 60  
DB 437 OSPTLTQSTNHTQSSSSSDGGLFRSPAHSLPPGEDGRVEPYVDFAEFRYRLMSYDHG 496

QY 61 EOSVVTAP 68  
DB 497 EOSVVTAP 504

RESULT 6  
US-09-406-854-6  
Sequence 6, Application US/09406854  
Patent No. 6140042  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihiko  
APPLICANT: NISHIDA, Eisuke  
TITLE OF INVENTION: TABI PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/406,854  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,891  
FILING DATE: 20-NOV-1996  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 17981/111  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 504 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-406-854-6

Query Match 100.0%; Score 359; DB 4; Length 504;  
Best Local Similarity 100.0%; Pred. No. 2.3e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 QSPFTLTQSTNTHTQSSSSDGLFRSRPAHSLPPGSDGRVPEYVDFAEFYRLWSVDHG 60  
Db 437 QSPFTLTQSTNTHTQSSSSDGLFRSRPAHSLPPGSDGRVPEYVDFAEFYRLWSVDHG 496  
Qy 61 EQSVVTAP 68  
Db 497 EQSVVTAP 504

RESULT 7  
US-09-529-279-2  
; Sequence 2, Application US/09529279  
; Patent No. 6451617  
; GENERAL INFORMATION:  
; APPLICANT: ONO, KOICHIRO  
; APPLICANT: OHTOMO, TOSHIHIKO  
; APPLICANT: TSUCHIYA, MASAYUKI  
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES  
; FILE REFERENCE: 053466/0278  
; CURRENT APPLICATION NUMBER: US/09/529, 279  
; CURRENT FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: PCT/JF98/04796  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: JP 9/290188  
; PRIOR FILING DATE: 1997-10-22  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 504  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-529-279-2

Query Match 100.0%; Score 359; DB 4; Length 504;  
Best Local Similarity 100.0%; Pred. No. 2.3e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 QSPFTLTQSTNTHTQSSSSDGLFRSRPAHSLPPGSDGRVPEYVDFAEFYRLWSVDHG 60  
Db 437 QSPFTLTQSTNTHTQSSSSDGLFRSRPAHSLPPGSDGRVPEYVDFAEFYRLWSVDHG 496  
Qy 61 EQSVVTAP 68  
Db 497 EQSVVTAP 504

RESULT 8  
US-09-529-279-43  
; Sequence 43, Application US/09529279  
; Patent No. 6451617  
; GENERAL INFORMATION:  
; APPLICANT: ONO, KOICHIRO  
; APPLICANT: OHTOMO, TOSHIHIKO  
; APPLICANT: TSUCHIYA, MASAYUKI  
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES  
; FILE REFERENCE: 053466/0278

; CURRENT APPLICATION NUMBER: US/09/529, 279  
; CURRENT FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: PCT/JF98/04796  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: JP 9/290188  
; PRIOR FILING DATE: 1997-10-22  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 43  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-529-279-43

Query Match 100.0%; Score 359; DB 4; Length 513;  
Best Local Similarity 100.0%; Pred. No. 2.4e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 QSPFTLTQSTNTHTQSSSSDGLFRSRPAHSLPPGSDGRVPEYVDFAEFYRLWSVDHG 60  
Db 446 QSPFTLTQSTNTHTQSSSSDGLFRSRPAHSLPPGSDGRVPEYVDFAEFYRLWSVDHG 505  
Qy 61 EQSVVTAP 68  
Db 506 EQSVVTAP 513

RESULT 9  
US-09-529-279-11  
; Sequence 11, Application US/09529279  
; Patent No. 6451617  
; GENERAL INFORMATION:  
; APPLICANT: ONO, KOICHIRO  
; APPLICANT: OHTOMO, TOSHIHIKO  
; APPLICANT: TSUCHIYA, MASAYUKI  
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES  
; FILE REFERENCE: 053466/0278  
; CURRENT APPLICATION NUMBER: US/09/529, 279  
; CURRENT FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: PCT/JF98/04796  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: JP 9/290188  
; PRIOR FILING DATE: 1997-10-22  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-529-279-11

Query Match 100.0%; Score 359; DB 4; Length 517;  
Best Local Similarity 100.0%; Pred. No. 2.4e-37;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 QSPFTLTQSTNTHTQSSSSDGLFRSRPAHSLPPGSDGRVPEYVDFAEFYRLWSVDHG 60  
Db 437 QSPFTLTQSTNTHTQSSSSDGLFRSRPAHSLPPGSDGRVPEYVDFAEFYRLWSVDHG 496  
Qy 61 EQSVVTAP 68  
Db 497 EQSVVTAP 504

RESULT 10  
US-09-529-279-41  
; Sequence 41, Application US/09529279  
; Patent No. 6451617  
; GENERAL INFORMATION:  
; APPLICANT: ONO, KOICHIRO  
; APPLICANT: OHTOMO, TOSHIHIKO  
; APPLICANT: TSUCHIYA, MASAYUKI  
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES

```

RESULT 13
US-09-529-279-40
Sequence 40, Application US/09529279
Patent No. 6451617
GENERAL INFORMATION:
APPLICANT: ONO, KOICHIRO
APPLICANT: OHNOMO, TOSHIOHKO
APPLICANT: TSUCHIYA, MASAYUKI
TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES
FILE REFERENCE: 053466/0278
CURRENT APPLICATION NUMBER: US/09/529,279
CURRENT FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: PCT/JF98/04796
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: JP 9/290188
PRIOR FILING DATE: 1997-10-22
NUMBER OF SEQ. ID NOS: 48
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 40
LENGTH: 16
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
OTHER INFORMATION: peptide

```

US-09-529-279-40

Query Match 18.4%; Score 66; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.032;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 VEPYVDFAEFYR 53

Db 2 VEPYVDFAEFYR 13

RESULT 14

US-09-152-060-100  
; Sequence 100, Application US/09152060  
; Patent No. 6448230  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 28 Human Secreted Proteins  
; FILE REFERENCE: P2003Pl US  
; CURRENT APPLICATION NUMBER: US/09/152,060  
; CURRENT FILING DATE: 1998-09-11  
; EARLIER APPLICATION NUMBER: PCT/US98/04858  
; EARLIER FILING DATE: 1998-03-12  
; EARLIER APPLICATION NUMBER: 60/040,762  
; EARLIER FILING DATE: 1997-03-14  
; EARLIER APPLICATION NUMBER: 60/040,710  
; EARLIER FILING DATE: 1997-03-14  
; EARLIER APPLICATION NUMBER: 60/050,934  
; EARLIER FILING DATE: 1997-05-30  
; EARLIER APPLICATION NUMBER: 60/048,100  
; EARLIER FILING DATE: 1997-05-30  
; EARLIER APPLICATION NUMBER: 60/048,357  
; EARLIER FILING DATE: 1997-05-30  
; EARLIER APPLICATION NUMBER: 60/048,189  
; EARLIER FILING DATE: 1997-05-30  
; EARLIER APPLICATION NUMBER: 60/057,765  
; EARLIER FILING DATE: 1997-09-05  
; EARLIER APPLICATION NUMBER: 60/048,970  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/068,368  
; EARLIER FILING DATE: 1997-12-19  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 100  
; LENGTH: 240  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-152-060-100

Query Match 17.7%; Score 63.5; DB 4; Length 240;  
Best Local Similarity 38.6%; Pred. No. 2.4;  
Matches 27; Conservative 2; Mismatches 30; Indels 11; Gaps 3;

QY 7 LQSTNTHQTSSSSSD-----GGLFRSRPAHSLPFGEDGRVEPY--VDFAEFYRLWS-- 56

Db 47 LQTAGTCGVSDSLGSLGPRTAAVFRPRPGDSLPSRTRYVELYVVVDNAEFQMLGSEA 106

QY 57 -VDHGEQSVV 65

Db 107 AVRHRVLEV 116

RESULT 15

US-09-152-060-64  
; Sequence 64, Application US/09152060  
; Patent No. 6448230  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 28 Human Secreted Proteins  
; FILE REFERENCE: P2003Pl US  
; CURRENT APPLICATION NUMBER: US/09/152,060  
; CURRENT FILING DATE: 1998-09-11  
; EARLIER APPLICATION NUMBER: PCT/US98/04858

; EARLIER FILING DATE: 1998-03-12  
; EARLIER APPLICATION NUMBER: 60/040,762  
; EARLIER FILING DATE: 1997-03-14  
; EARLIER APPLICATION NUMBER: 60/040,710  
; EARLIER FILING DATE: 1997-03-14  
; EARLIER APPLICATION NUMBER: 60/050,934  
; EARLIER FILING DATE: 1997-05-30  
; EARLIER APPLICATION NUMBER: 60/048,100  
; EARLIER FILING DATE: 1997-05-30  
; EARLIER APPLICATION NUMBER: 60/048,357  
; EARLIER FILING DATE: 1997-05-30  
; EARLIER APPLICATION NUMBER: 60/048,189  
; EARLIER FILING DATE: 1997-05-30  
; EARLIER APPLICATION NUMBER: 60/057,765  
; EARLIER FILING DATE: 1997-09-05  
; EARLIER APPLICATION NUMBER: 60/048,970  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/068,368  
; EARLIER FILING DATE: 1997-12-19  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 64  
; LENGTH: 335  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (35)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (297)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-152-060-64

Query Match 17.7%; Score 63.5; DB 4; Length 335;  
Best Local Similarity 38.6%; Pred. No. 3.7;  
Matches 27; Conservative 2; Mismatches 30; Indels 11; Gaps 3;

QY 7 LQSTNTHQTSSSSSD-----GGLFRSRPAHSLPFGEDGRVEPY--VDFAEFYRLWS-- 56

Db 161 LQTAGTCGVSDSLGSLGPRTAAVFRPRPGDSLPSRTRYVELYVVVDNAEFQMLGSEA 220

QY 57 -VDHGEQSVV 65

Db 221 AVRHRVLEV 230

Search completed: December 10, 2002, 07:35:24  
Job time : 40 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 07:30:46 ; Search time 70 Seconds  
(Without alignments)  
15.778 Million cell updates/sec

Title: US-09-830-144-4\_COPY\_437\_504

Perfect score: 359  
Sequence: 1 OSPTLTIGSTNTHHTQSSSSS.....AEFYRLMSVDHGSGVVTAP 68

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 103943 seqs, 16242309 residues

Total number of hits satisfying chosen parameters: 103943

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PTC\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PTCUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	359	100.0	504	9 US-10-158-895-2	Sequence 2, Appli
2	359	100.0	504	12 US-10-123-427-2	Sequence 2, Appli
3	359	100.0	504	12 US-10-123-427-6	Sequence 6, Appli
4	359	100.0	513	9 US-10-158-895-43	Sequence 43, Appli
5	359	100.0	517	9 US-10-158-895-11	Sequence 11, Appli
6	352	98.1	84	10 US-09-925-300-1270	Sequence 1270, Ap
7	77	21.4	16	9 US-10-158-895-41	Sequence 41, Appli
8	69.5	19.4	199	9 US-09-941-831-21	Sequence 21, Appli
9	66	18.4	16	9 US-10-158-895-40	Sequence 40, Appli
10	65.5	18.2	498	9 US-10-037-667-5	Sequence 5, Appli
11	63.5	17.7	240	9 US-09-852-797-100	Sequence 100, App
12	63.5	17.7	240	10 US-09-853-161-100	Sequence 100, App
13	63.5	17.7	240	10 US-09-852-659A-100	Sequence 100, App
14	63.5	17.7	335	9 US-09-852-797-64	Sequence 64, Appli
15	63.5	17.7	335	10 US-09-853-161-64	Sequence 64, Appli
16	63.5	17.7	335	10 US-09-852-659A-64	Sequence 64, Appli
17	63	17.5	850	10 US-09-915-181A-3	Sequence 3, Appli
18	61	17.0	322	10 US-09-764-853-606	Sequence 606, App
19	61	17.0	1511	10 US-09-801-368-250	Sequence 250, App

20	58.5	16.3	235	10 US-09-799-777-43	Sequence 43, Appli
21	58.5	16.3	3256	10 US-09-919-172-98	Sequence 98, Appli
22	57.5	16.0	394	10 US-09-815-242-10233	Sequence 10233, A
23	57.5	16.0	421	10 US-09-965-553-6	Sequence 6, Appli
24	57.5	16.0	856	9 US-09-852-797-77	Sequence 77, Appli
25	57.5	16.0	856	10 US-09-853-161-77	Sequence 77, Appli
26	57.5	16.0	856	10 US-09-852-659A-77	Sequence 1643, Ap
27	56.5	15.7	255	10 US-09-764-877-1843	Sequence 2, Appli
28	56.5	15.7	923	12 US-10-104-440-2	Sequence 26, Appli
29	56	15.6	427	10 US-08-681-219-26	Sequence 5, Appli
30	56	15.6	427	10 US-09-826-212-5	Sequence 13, Appli
31	56	15.6	427	10 US-09-748-537-13	Sequence 7, Appli
32	56	15.6	427	10 US-09-935-727-7	Sequence 4, Appli
33	56	15.6	427	10 US-09-756-854-4	Sequence 4, Appli
34	56	15.6	455	9 US-10-041-574-4	Sequence 5824, Ap
35	56	15.6	887	10 US-09-815-242-5824	Sequence 12997, A
36	56	15.6	887	10 US-09-815-242-12997	Sequence 36769, A
37	55	15.3	125	10 US-09-864-761-36769	Sequence 4, Appli
38	55	15.3	822	9 US-10-003-295-4	Sequence 46377, A
39	54.5	15.2	338	10 US-09-864-761-46977	Sequence 69, Appli
40	54	15.0	1463	9 US-09-971-536-69	Sequence 36662, A
41	53.5	14.9	162	10 US-09-864-761-36662	Sequence 45403, A
42	53.5	14.9	162	10 US-09-864-761-45403	Sequence 2, Appli
43	53	14.8	760	10 US-09-265-606-2	Sequence 67, Appli
44	53	14.8	783	10 US-09-888-615-67	Sequence 5, Appli
45	53	14.8	814	10 US-09-808-568-5	

#### ALIGNMENTS

RESULT 1  
US-10-158-895-2  
; Sequence 2, Application US/10158895  
; Patent No. US20020155624A1  
; GENERAL INFORMATION:  
; APPLICANT: ONO, KOICHIRO  
; APPLICANT: OHTOMO, TOSHIHIKO  
; APPLICANT: TSUCHIYA, MASAYUKI  
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES  
; FILE REFERENCE: 053466/0278  
; CURRENT APPLICATION NUMBER: US/10/158, 895  
; CURRENT FILING DATE: 2002-06-03  
; PRIOR APPLICATION NUMBER: US/09/529, 279  
; PRIOR FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: PCT/JP98/04796  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: JP 9/290188  
; PRIOR FILING DATE: 1997-10-22  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 504  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-158-895-2

Query Match 100.0%; Score 359; DB 9; Length 504;

Best Local Similarity 100.0%; Pred. No. 1.1e-34;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 OSPTLTIGSTNTHHTQSSSSSDGGLFRSRPAHSLPRGSDGRVEPYVDFAEFYRLMSVDHG 60  
DB 437 OSPTLTIGSTNTHHTQSSSSSDGGLFRSRPAHSLPRGSDGRVEPYVDFAEFYRLMSVDHG 496  
QY 61 EOSVVTAP 68  
DB 497 EOSVVTAP 504

RESULT 2  
US-10-123-427-2  
; Sequence 2, Application US/10123427

Patent No. US20020119525A1  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihiro  
NISHIDA, Eisuke  
TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/123,427  
FILING DATE: 17-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/406,854  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US/08/752,891  
FILING DATE: 20-NOV-1996  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 504 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-10-123-427-2

Query Match 100.0%; Score 359; DB 12; Length 504;  
Best Local Similarity 100.0%; Pred. No. 1.1e-34;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QSPFTLTQSTNTHQTSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDFAEFYRLWSVDHG 60  
Db 437 QSPFTLTQSTNTHQTSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDFAEFYRLWSVDHG 496

Qy 61 EQSVVTAP 68  
Db 497 EQSVVTAP 504

RESULT 3  
US-10-123-427-6  
Sequence 6, Application US/10123427  
Patent No. US20020119525A1  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihiro  
NISHIDA, Eisuke  
TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500

CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/123,427  
FILING DATE: 17-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/406,854  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US/08/752,891  
FILING DATE: 20-NOV-1996  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 504 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-123-427-6

Query Match 100.0%; Score 359; DB 12; Length 504;  
Best Local Similarity 100.0%; Pred. No. 1.1e-34;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QSPFTLTQSTNTHQTSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDFAEFYRLWSVDHG 60  
Db 437 QSPFTLTQSTNTHQTSSSSSDGGLFRSRPAHSLPPGEDGRVEPYVDFAEFYRLWSVDHG 496

Qy 61 EQSVVTAP 68  
Db 497 EQSVVTAP 504

RESULT 4  
US-10-158-895-43  
Sequence 43, Application US/10158895  
Patent No. US20020155624A1  
GENERAL INFORMATION:  
APPLICANT: ONO, KOICHIRO  
APPLICANT: OHTOMO, TOSHIHIKO  
APPLICANT: TSUCHIYA, MASAYUKI  
TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES  
FILE REFERENCE: 053466/0278  
CURRENT APPLICATION NUMBER: US/10/158,895  
CURRENT FILING DATE: 2002-06-03  
PRIOR APPLICATION NUMBER: US/09/529,279  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: PCT/JP98/04796  
PRIOR FILING DATE: 1998-10-22  
PRIOR APPLICATION NUMBER: JP 9/290188  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 43



```
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-158-895-43

Query Match          100.0%; Score 359; DB 9; Length 513;
Best Local Similarity 100.0%; Pred. No. 1,1e-34;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 OSPITLLOSTNTHTQSSSSSDGSLFRSRPAHSLPGEEDGRVEPYVDAEFYRLMSVDHG 60
   |||||
Db 446 OSPITLLOSTNTHTQSSSSSDGSLFRSRPAHSLPGEEDGRVEPYVDAEFYRLMSVDHG 505
   |||||

QY 61 EQSVVTAP 68
   |||||
Db 506 EQSVVTAP 513
   |||||

RESULT 5
US-10-158-895-11
; Sequence 11, Application US/10158895
; Patent No. US20020155624A1
; GENERAL INFORMATION:
; APPLICANT: ONO, KOICHIRO
; APPLICANT: OHTOMO, TOSHIHIKO
; APPLICANT: TSUCHIYA, MASAYUKI
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES
; FILE REFERENCE: 053466/0278
; CURRENT APPLICATION NUMBER: US/10/158, 895
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/529, 279
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: PCT/JP98/04796
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: JP 9/290188
; PRIOR FILING DATE: 1997-10-22
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-158-895-11

Query Match          100.0%; Score 359; DB 9; Length 517;
Best Local Similarity 100.0%; Pred. No. 1,1e-34;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 OSPITLLOSTNTHTQSSSSSDGSLFRSRPAHSLPGEEDGRVEPYVDAEFYRLMSVDHG 60
   |||||
Db 437 OSPITLLOSTNTHTQSSSSSDGSLFRSRPAHSLPGEEDGRVEPYVDAEFYRLMSVDHG 496
   |||||

QY 61 EQSVVTAP 68
   |||||
Db 497 EQSVVTAP 504
   |||||

RESULT 6
US-09-925-300-1270
; Sequence 1270, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925, 300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1270
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (38)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1270

Query Match          98.1%; Score 352; DB 10; Length 84;
Best Local Similarity 98.5%; Pred. No. 8.9e-35;
Matches 67; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 OSPITLLOSTNTHTQSSSSSDGSLFRSRPAHSLPGEEDGRVEPYVDAEFYRLMSVDHG 60
   |||||
Db 17 OSPITLLOSTNTHTQSSSSSDGSLFRSRPAHSLPGEEDGRVEPYVDAEFYRLMSVDHG 76
   |||||

QY 61 EQSVVTAP 68
   |||||
Db 77 EQSVVTAP 84
   |||||

RESULT 7
US-10-158-895-41
; Sequence 41, Application US/10158895
; Patent No. US20020155624A1
; GENERAL INFORMATION:
; APPLICANT: ONO, KOICHIRO
; APPLICANT: OHTOMO, TOSHIHIKO
; APPLICANT: TSUCHIYA, MASAYUKI
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES
; FILE REFERENCE: 053466/0278
; CURRENT APPLICATION NUMBER: US/10/158, 895
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/529, 279
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: PCT/JP98/04796
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: JP 9/290188
; PRIOR FILING DATE: 1997-10-22
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-158-895-41

Query Match          21.4%; Score 77; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 OSPITLLOSTNTHTQ 15
   |||||
Db 2 OSPITLLOSTNTHTQ 16
   |||||

RESULT 8
US-09-941-831-21
; Sequence 21, Application US/09941831
; Patent No. US20020160493A1
; GENERAL INFORMATION:
; APPLICANT: Ebner et al.
; TITLE OF INVENTION: PT049P1
; FILE REFERENCE: Serine/threonine Phosphatase Polynucleotides, Polypeptides, and Antib
; CURRENT APPLICATION NUMBER: US/09/941, 831
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: PCT/US01/06256
```

; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 60/186,350  
; PRIOR FILING DATE: 2000-03-02  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 21  
; LENGTH: 199  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-941-831-21

Query Match 19.4%; Score 69.5; DB 9; Length 199;  
Best Local Similarity 41.9%; Pred. No. 0.36;  
Matches 18; Conservative 3; Mismatches 17; Indels 5; Gaps 1;

QY 1 QSPFTLQSTNTHTQSSSSDGG-----LFRSRPAHSLPGE 38  
Db 119 QSPFTLQDQTRIHSDRAFSISGCSKFTAVRRKRWADKLPVQ 161

RESULT 9

US-10-158-895-40  
; Sequence 40, Application US/10158895  
; Patent No. US2002015562A1

; GENERAL INFORMATION:  
; APPLICANT: ONO, KOICHIRO  
; APPLICANT: TOSHOMO, TOSHIHIKO  
; APPLICANT: TSUCHIYA, MASAYUKI  
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES  
; FILE REFERENCE: 053466/0278  
; CURRENT APPLICATION NUMBER: US/10/158,895  
; CURRENT FILING DATE: 2002-06-03  
; PRIOR APPLICATION NUMBER: US/09/529,279  
; PRIOR FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: PCT/JP98/04796  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: JP 9/290188  
; PRIOR FILING DATE: 1997-10-22  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 40  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
US-10-158-895-40

Query Match 18.4%; Score 66; DB 9; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 VEPYVDFAEFYR 53  
Db 2 VEPYVDFAEFYR 13

RESULT 10

US-10-037-667-5  
; Sequence 5, Application US/10037667  
; Patent No. US2002017145A1

; GENERAL INFORMATION:  
; APPLICANT: Morgan, Bruce A.  
; TITLE OF INVENTION: REGULATION OF NEURAL DEVELOPMENT BY  
; FILE REFERENCE: DAEDALOS  
; CURRENT APPLICATION NUMBER: US/10/037,667  
; CURRENT FILING DATE: 2002-07-23  
; PRIOR APPLICATION NUMBER: 60/243,110  
; PRIOR FILING DATE: 2000-10-25  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5  
; LENGTH: 498  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-037-667-5

Query Match 18.2%; Score 65.5; DB 9; Length 498;  
Best Local Similarity 31.8%; Pred. No. 3.1;  
Matches 28; Conservative 9; Mismatches 18; Indels 33; Gaps 5;

QY 2 SPTLTQSTNTHTQSS-----SDGGLFRSRPAHSLPP--- 36  
Db 372 SPTNGQDSNTDTESNHEERGQATSSRSQSSAYAKEQRPDSGGL--LFRSMPTAKE 429  
QY 37 -----CEDGRVPEPYVDFAEFYRLMSVDH 59  
Db 430 SLRVLGEDG-VQVKVFKCEHCRVFLDH 456

RESULT 11

US-09-852-797-100  
; Sequence 100, Application US/09852797  
; Patent No. US20020172994A1

; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 28 Human Secreted Proteins  
; FILE REFERENCE: P2003P2  
; CURRENT APPLICATION NUMBER: US/09/852,797  
; CURRENT FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: 60/265,583  
; PRIOR FILING DATE: 2001-02-02  
; PRIOR APPLICATION NUMBER: 09/152,060  
; PRIOR FILING DATE: 1998-09-11  
; PRIOR APPLICATION NUMBER: PCT/US98/04858  
; PRIOR FILING DATE: 1998-03-12  
; PRIOR APPLICATION NUMBER: 60/040,762  
; PRIOR FILING DATE: 1997-03-14  
; PRIOR APPLICATION NUMBER: 60/040,710  
; PRIOR FILING DATE: 1997-03-14  
; PRIOR APPLICATION NUMBER: 60/050,934  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: 60/048,100  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: 60/048,357  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: 60/048,189  
; PRIOR FILING DATE: 1997-05-30  
; PRIOR APPLICATION NUMBER: 60/057,765  
; PRIOR FILING DATE: 1997-09-05  
; PRIOR APPLICATION NUMBER: 60/048,970  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/068,368  
; PRIOR FILING DATE: 1997-12-19  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 100  
; LENGTH: 240  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-852-797-100

Query Match 17.7%; Score 63.5; DB 9; Length 240;  
Best Local Similarity 38.6%; Pred. No. 2.3;  
Matches 27; Conservative 2; Mismatches 30; Indels 11; Gaps 3;

QY 7 LQSTNTHQSSSSSD-----GGLFRSRPAHSLPPGEDGRVEPY--VDFAEFYRLWS-- 56  
Db 47 LQTAGTCGVSDSDSLGSLGPRTAATAVFRPRPGDSLPSRETRYVELYVVVDNAEFQMLGSEA 106  
QY 57 -VDHGEQSVV 65  
Db 107 AVNRHRLVVV 116

```

RESULT 12
US-09-853-161-100
; Sequence 100, Application US/09853161
; Patent No. US20020076756A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
FILE REFERENCE: P2003P3
CURRENT APPLICATION NUMBER: US/09/853,161
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: 60/265,583
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 09/152,060
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: PCT/US98/04858
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/040,762
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/040,710
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/050,934
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,100
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,357
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,189
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/057,765
PRIOR FILING DATE: 1997-09-05
PRIOR APPLICATION NUMBER: 60/048,970
PRIOR FILING DATE: 1997-06-06
PRIOR APPLICATION NUMBER: 60/068,368
PRIOR FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 118
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 100
LENGTH: 240
TYPE: PRT
ORGANISM: Homo sapiens
US-09-853-161-100

Query Match          17.7%; Score 63.5; DB 10; Length 240;
Best Local Similarity 38.6%; Pred. No. 2.3;
Matches 27; Conservative 2; Mismatches 30; Indels 11; Gaps 3;

QY      7 LGSTNTHGSSSSSD-----GGIFRSRPAHSLPPGEDGRVPEY--VDFAFFYLMS-- 56
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db       47 LQTAGCGVSDSLGLSLGPRTAAVFRPRPDLSLPERETRYELVVVDMAEFOMLGSEA 106

QY      57 -VDHGQSIVV 65
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db       107 AVHRVLELV 116

RESULT 13
US-09-852-659A-100
; Sequence 100, Application US/09852659A
; Patent No. US20020077287A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
FILE REFERENCE: P2003P4
CURRENT APPLICATION NUMBER: US/09/852,659A
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: 60/265,583
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 09/152,060
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: PCT/US98/04858
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/040,762
```

```

PQ      7  LOSTTHHQSSSSSD-----GGLFRSRPHSLPPGEDGVPEY--VDFAEFYRLWS-- 56
OY      47  LQTACTCGVSDSLSGLSLGPRTAAVFRRFPDGSUPRSRETRYVELVWVDNMEFOWLGSEA 106
DB      107 AVRHRLEV 116

US-09-852-659A-100
ORGANISM: Homo sapiens
TYPE: PRT
LENGTH: 240
SEQ ID NO 100
SOFTWARE: PatentIn Ver. 2.0

Prior Filing Date: 1997-03-14
Prior Application Number: 60/040,710
Prior Filing Date: 1997-03-14
Prior Application Number: 60/050,934
Prior Filing Date: 1997-05-30
Prior Application Number: 60/048,100
Prior Filing Date: 1997-05-30
Prior Application Number: 60/048,357
Prior Filing Date: 1997-05-30
Prior Application Number: 60/048,189
Prior Filing Date: 1997-05-30
Prior Application Number: 60/057,765
Prior Filing Date: 1997-09-05
Prior Application Number: 60/048,970
Prior Filing Date: 1997-06-06
Prior Application Number: 60/068,368
Prior Filing Date: 1997-12-19
NUMBER OF SEQ ID NOS: 121

US-09-852-659A-100
ORGANISM: Homo sapiens
TYPE: PRT
LENGTH: 240
SEQ ID NO 100
SOFTWARE: PatentIn Ver. 2.0

Query Match          17.7%; Score 63.5; DB 10; Length 240;
Best Local Similarity 38.6%; Pred. No. 2.3;
Matches 27; Conservative 2; Mismatches 30; Indels 11; Gaps 3.

OY      7  LOSTTHHQSSSSSD-----GGLFRSRPHSLPPGEDGVPEY--VDFAEFYRLWS-- 56
DB      47  LQTACTCGVSDSLSGLSLGPRTAAVFRRFPDGSUPRSRETRYVELVWVDNMEFOWLGSEA 106
DB      107 AVRHRLEV 116

RESULT 14
US-09-852-797-64
Sequence 64, Application US/09852797
Patent No. US20020172994A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.,
TITLE OF INVENTION: 28 Human Secreted Proteins
FILE REFERENCE: PZ003P2
CURRENT APPLICATION NUMBER: US/09/852,797
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: 60/265,583
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 09/152,060
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: PCT/US98/04858
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/040,762
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/040,710
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/050,934
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,100
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,357
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,189
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/057,765
PRIOR FILING DATE: 1997-09-05
PRIOR APPLICATION NUMBER: 60/048,970
PRIOR FILING DATE: 1997-06-06
PRIOR APPLICATION NUMBER: 60/068,368
PRIOR FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 118
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (35)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (297)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-852-797-64

Query Match      17.7%; Score 63.5; DB 9; Length 335;
Best Local Similarity 38.6%; Pred. No. 3.3;
Matches 27; Conservative 2; Mismatches 30; Indels 11; Gaps 3;

QY 7 LQSTNTHQTQSSSSSD-----GGLFRSRPAHSLPPGEDGRVEPY--VDFAEFYRLMS-- 56
Db 161 LQTAGTCGVSDSLGSLGPRTAAVFRPRPGDSLPSRETRYVELYVVVDNAEFQMLGSEA 220

QY 57 -VDHGEQSVV 65
Db 221 AVRHRVLEVV 230

Search completed: December 10, 2002, 07:36:46
Job time : 70 secs

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-853-161-64

Query Match      17.7%; Score 63.5; DB 10; Length 335;
Best Local Similarity 38.6%; Pred. No. 3.3;
Matches 27; Conservative 2; Mismatches 30; Indels 11; Gaps 3;

QY 7 LQSTNTHQTQSSSSSD-----GGLFRSRPAHSLPPGEDGRVEPY--VDFAEFYRLMS-- 56
Db 161 LQTAGTCGVSDSLGSLGPRTAAVFRPRPGDSLPSRETRYVELYVVVDNAEFQMLGSEA 220

QY 57 -VDHGEQSVV 65
Db 221 AVRHRVLEVV 230

Search completed: December 10, 2002, 07:36:46
Job time : 70 secs

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (35)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (297)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-852-797-64

Query Match      17.7%; Score 63.5; DB 9; Length 335;
Best Local Similarity 38.6%; Pred. No. 3.3;
Matches 27; Conservative 2; Mismatches 30; Indels 11; Gaps 3;

QY 7 LQSTNTHQTQSSSSSD-----GGLFRSRPAHSLPPGEDGRVEPY--VDFAEFYRLMS-- 56
Db 161 LQTAGTCGVSDSLGSLGPRTAAVFRPRPGDSLPSRETRYVELYVVVDNAEFQMLGSEA 220

QY 57 -VDHGEQSVV 65
Db 221 AVRHRVLEVV 230

RESULT 15
US-09-853-161-64
; Sequence 64, Application US/09853161
; Patent No. US20020076756A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P3
; CURRENT APPLICATION NUMBER: US/09/853,161
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (35)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (297)
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: December 10, 2002, 04:57:56 ; Search time 3101 Seconds  
(without alignments)  
638.178 Million cell updates/sec

Title: US-09-830-144-4\_COPY\_437\_504

Perfect score: 359 1 QSFPTLTGSTNTHHTGSSSSS.....AEFTRLMSVDHGOSVVTAP 68

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODL=frame+g2n.model -DEV=x1p  
-O=/cgm2\_1/USPTO.spool/US09830144/runat\_04122002\_141842\_4974/app\_query\_fastq\_1.263  
-DB=genembl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bitsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09830144 @CGN 1 1 3637 @runat\_04122002\_141842\_4974 -NCPUs=6 -ICPU=3  
-NO\_XIPYX -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hlg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_hlg\_hum.\*  
31: em\_hlg\_inv.\*  
32: em\_hlg\_other.\*  
33: em\_hlg\_mus.\*  
34: em\_hlg\_pln.\*  
35: em\_hlg\_rod.\*  
36: em\_hlg\_mam.\*  
37: em\_hlg\_vrt.\*  
38: em\_sy.\*  
39: em\_higo\_hum.\*  
40: em\_higo\_mus.\*  
41: em\_higo\_other.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	359	100.0	1535	6 E31041	E31041 Method for
2	359	100.0	1560	6 AR058299	AR058299 Sequence
3	359	100.0	1560	6 AR058302	AR058302 Sequence
4	359	100.0	1560	6 AR088273	AR088273 Sequence
5	359	100.0	1560	6 AR088276	AR088276 Sequence
6	359	100.0	1560	6 AR116881	AR116881 Sequence
7	359	100.0	1560	6 AR116884	AR116884 Sequence
8	359	100.0	1560	6 E14752	E14752 Human mRNA
9	359	100.0	3096	6 HS049928	U49928 Homo sapien
10	359	100.0	69660	9 HS047054	Z83845 Human DNA s
11	356	99.2	2944	10 BC027054	BC027054 Mus muscu
12	356	99.2	141498	2 AC127924	AC127924 Rattus no
13	356	99.2	16665	2 AC127784	AC127784 Rattus no
14	260.5	92.6	1926	5 XLU92031	U92031 Xenopus lae
15	79	22.0	24526	2 ALB45323	ALB45323 Mus muscu
16	78	21.7	153463	2 AC115155	AC115155 Rattus no
17	76	21.2	519	6 AX386178	AX386178 Sequence
18	76	21.2	113028	9 AL365220	AL365220 Human DNA
19	75.5	21.0	82166	9 AC006237	AC006237 Homo sapi
20	75.5	21.0	179966	2 AC023247	AC023247 Homo sapi
21	75.5	21.0	199734	2 AC102141	AC102141 Mus muscu
22	75.5	21.0	205040	9 AC015849	AC015849 Homo sapi
23	75.5	21.0	241541	2 AC084085	AC084085 Homo sapi
24	75	20.9	14090	2 AC014059	AC014059 Drosophi1
25	75	20.9	158402	3 AC008343	AC008343 Drosophi1
26	75	20.9	25652	3 AE003811	AE003811 Drosophi1
27	74	20.6	124271	2 AC123505	AC123505 Rattus no
28	74	20.6	164521	2 AC119354	AC119354 Rattus no
29	74	20.6	165265	2 AC117064	AC117064 Rattus no
30	74	20.6	171111	2 AC099452	AC099452 Rattus no
31	74	20.6	223761	2 AC127239	AC127239 Mus muscu
32	74	20.6	229247	10 AC098888	AC098888 Mus muscu
33	74	20.6	317515	2 AC099415	AC099415 Mus muscu
34	73.5	20.5	169264	2 AC112889	AC112889 Rattus no
35	73	20.3	3438	1 ABU55611	U35611 Azospirillum
36	73	20.3	4440	1 AFI50957	AFI50957 Azospirill
37	73	20.3	34266	3 LMFJ391	ALJ35941 Leishmani
38	73	20.3	139130	9 AC091504	AC091504 Pan trogl
39	73	20.3	163849	2 AC044808	AC044808 Mus muscu
40	73	20.3	226048	2 AL669881	AL669881 Mus muscu
41	72.5	20.2	131888	9 HSJ10574	ALJ121777 Human DNA
42	72.5	20.2	144877	2 AC128571	AC128571 Rattus no
43	72	20.1	10029	1 AE013381	AE013381 Methanosa
44	72	20.1	77796	2 AC068176	AC068176 Homo sapi
45	72	20.1	90732	9 AL670886	AL670886 Human DNA

RESULT 1

#### ALIGNMENTS

E31041  
LOCUS E31041 1515 bp DNA linear PAT 18-JUN-2001  
DEFINITION Method for screening substance inhibiting binding to XIAP.  
ACCESSION E31041  
VERSION E31041.1 GI:13017306  
KEYWORDS JP 1999326328-A/1.  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1515)  
AUTHORS Kunihiro, M.  
TITLE Method for screening substance inhibiting binding to XIAP  
JOURNAL KUNIHIRO MATSUMOTO  
COMMENT OS Unidentified  
PN JP 1999326328-A/1  
PD 26-NOV-1999  
PF 13-MAY-1998 JP 1998130378  
PR KUNIHIRO MATSUMOTO  
PI  
PC GOIN33/566, A61K38/00, A61K38/00, A61K38/00, A61K38/22,  
PC A61K39/395,  
PC A61K45/00, A61K45/00, C07K7/06, C07K7/08, C07K14/47, PC  
GOIN33/535,  
PC GOIN33/535, GOIN33/536, C12N15/09, C12P21/08, A61K37/02,  
A61K37/02, PC A61K37/02,  
PC A61K37/02, A61K37/24, C12N15/00  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT CDS 1..1512.  
          Location/Qualifiers  
          1..1515  
          /organism="unidentified"  
          /db\_xref="taxon:32644"  
BASE COUNT 323 a 457 c 463 g 272 t  
ORIGIN  
  
Alignment Scores:  
Pred. No.: 1,98e-30 Length: 1515  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
  
US-09-830-144-4\_COPY\_437\_504 (1-68) x E31041 (1-1515)  
  
QY 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
Db 1309 CAAAGCCCGACCTTAACCTCGAGTCCACCAACACGACGACGACGACGACGACG 1368  
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40  
Db 1369 TCTGACGGAGGCTCTTCGCTCCCGCCGCCACCTCGCTCCGCTCGGAGGACGGT 1428  
QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
Db 1429 CGTGTGAGCCCTATGTGACCTTCTGAGTTTACCGCTCTCGGAGGACGACGACG 1488  
QY 61 GluGlnSerValValThrAlaPro 68  
Db 1489 GAGCAGAGCGTGGTGACAGCACC 1512  
  
RESULT 2  
LOCUS AR058299 1560 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 1 from patent US 5837819.  
ACCESSION AR058299  
VERSION AR058299.1 GI:5983876  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

Unclassified.  
REFERENCE 1 (bases 1 to 1560)  
AUTHORS Matsumoto, K. and Nishida, E.  
TITLE TAB1 protein  
JOURNAL Patent: US 5837819-A 1 17-NOV-1998;  
FEATURES Location/Qualifiers  
          source 1..1560  
          BASE COUNT 332 a 469 c 480 g 279 t  
          ORIGIN  
  
Alignment Scores:  
Pred. No.: 2,04e-30 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
  
US-09-830-144-4\_COPY\_437\_504 (1-68) x AR058299 (1-1560)  
  
QY 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
Db 1338 CAAAGCCCGACCTTAACCTCGAGTCCACCAACACGACGACGACGACGACGACG 1397  
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40  
Db 1398 TCTGACGGAGGCTCTTCGCTCCCGCCGCCACCTCGCTCCGCTCGGAGGACGGT 1457  
QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
Db 1458 CGTGTGAGCCCTATGTGACCTTCTGAGTTTACCGCTCTCGGAGGACGACGACG 1517  
QY 61 GluGlnSerValValThrAlaPro 68  
Db 1518 GAGCAGAGCGTGGTGACAGCACC 1541  
  
RESULT 3  
LOCUS AR058302 1560 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 5 from patent US 5837819.  
ACCESSION AR058302  
VERSION AR058302.1 GI:5983879  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 1560)  
AUTHORS Matsumoto, K. and Nishida, E.  
TITLE TAB1 protein  
JOURNAL Patent: US 5837819-A 5 17-NOV-1998;  
FEATURES Location/Qualifiers  
          source 1..1560  
          BASE COUNT 333 a 468 c 480 g 279 t  
          ORIGIN  
  
Alignment Scores:  
Pred. No.: 2,04e-30 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
  
US-09-830-144-4\_COPY\_437\_504 (1-68) x AR058302 (1-1560)  
  
QY 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
Db 1338 CAAAGCCCGACCTTAACCTCGAGTCCACCAACACGACGACGACGACGACGACG 1397  
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40

Db 1398 TCTGACGAGGCTCTTCCGCTCCGCGCCGCACTCGCTCCGCTGCGAGGACGCT 1457  
Qy 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
Db 1458 CGTGTGAGCCCTTAAGTGAAGCTTTGAGAGTTTACCGCTCGAGCGTGAACCATGAC 1517  
Qy 61 GluGlnSerValValThrAlaPro 68  
Db 1518 GAGCAGACGCTGTGACGACCG 1541

RESULT 4  
AR088273  
LOCUS AR088273 1560 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 1 from patent US 5989862.  
ACCESSION AR088273.1 GI:10015036  
VERSION AR088273.1 GI:10015036  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1560)  
AUTHORS Matsumoto,K. and Nishida,E.  
TITLE Tab1 protein and DNA coding therefor  
JOURNAL Patent: US 5989862-A 1 23-NOV-1999;  
FEATURES  
source location/Qualifiers  
1..1560  
/organism="unknown"

BASE COUNT 332 a 469 c 480 g 279 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2.04e-30 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x AR088273 (1-1560)

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
Db 1338 CAAAGCCCGACCTTAACCTTGAGCTTGCAGATTTCACGACGACGAGGAGCTTCAGC 1397  
Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGly 40  
Db 1398 TCTGACGAGGCTCTTCCGCTCCGCGCCGCACTCGCTCCGCTGCGAGGACGCT 1457  
Qy 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
Db 1458 CGTGTGAGCCCTTAAGTGAAGCTTTGAGAGTTTACCGCTCGAGCGTGAACCATGAC 1517  
Qy 61 GluGlnSerValValThrAlaPro 68  
Db 1518 GAGCAGACGCTGTGACGACCG 1541

RESULT 5  
AR088276  
LOCUS AR088276 1560 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 5 from patent US 5989862.  
ACCESSION AR088276  
VERSION AR088276.1 GI:10015039  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1560)  
AUTHORS Matsumoto,K. and Nishida,E.  
TITLE Tab1 protein and DNA coding therefor  
JOURNAL Patent: US 5989862-A 5 23-NOV-1999;  
FEATURES  
source location/Qualifiers  
1..1560  
/organism="unknown"

BASE COUNT 333 a 468 c 480 g 279 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2.04e-30 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x AR088276 (1-1560)

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
Db 1338 CAAAGCCCGACCTTAACCTTGAGCTTGCAGATTTCACGACGACGAGGAGCTTCAGC 1397  
Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGly 40  
Db 1398 TCTGACGAGGCTCTTCCGCTCCGCGCCGCACTCGCTCCGCTGCGAGGACGCT 1457  
Qy 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
Db 1458 CGTGTGAGCCCTTAAGTGAAGCTTTGAGAGTTTACCGCTCGAGCGTGAACCATGAC 1517  
Qy 61 GluGlnSerValValThrAlaPro 68  
Db 1518 GAGCAGACGCTGTGACGACCG 1541

RESULT 6  
AR116881  
LOCUS AR116881 1560 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 1 from patent US 6140042.  
ACCESSION AR116881  
VERSION AR116881.1 GI:14097787  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1560)  
AUTHORS Matsumoto,K. and Nishida,E.  
TITLE TAB1 protein and DNA coding therefor  
JOURNAL Patent: US 6140042-A 1 31-OCT-2000;  
FEATURES  
source location/Qualifiers  
1..1560  
/organism="unknown"

BASE COUNT 332 a 469 c 480 g 279 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2.04e-30 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x AR116881 (1-1560)

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
Db 1338 CAAAGCCCGACCTTAACCTTGAGCTTGCAGATTTCACGACGACGAGGAGCTTCAGC 1397  
Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGly 40  
Db 1398 TCTGACGAGGCTCTTCCGCTCCGCGCCGCACTCGCTCCGCTGCGAGGACGCT 1457  
Qy 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
Db 1458 CGTGTGAGCCCTTAAGTGAAGCTTTGAGAGTTTACCGCTCGAGCGTGAACCATGAC 1517  
Qy 61 GluGlnSerValValThrAlaPro 68  
Db 1518 GAGCAGACGCTGTGACGACCG 1541

PC	(C12N5/10,
PC	C12R1:91),(C12P21/02,C12R1:865),(C12P21/02,C12R1:91); CC
strandedness:	Double;
CC	topology: Linear;
PH	Key
PH	Location/Qualifiers
FT	source
FT	1..1560
FT	/organism='Homo sapiens'
FT	/tissue_type='kidney'
FT	30..1544
CDS	
FT	/product='the TAB1 protein'
FT	185
FT	/note='replace a'
FEATURES	
source	Location/Qualifiers
1..1560	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
332 a	469 c 480 g 279 t
BASE COUNT	
ORIGIN	
Alignment Scores:	
Pred. No.:	2.04e-30 Length: 1560
Score:	359.00 Matches: 68
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB:	6 Gaps: 0
US-09-830-144-4_COPY_437_504 (1-68) x E14752 (1-1560)	
QY	1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20
Db	1338 CAAGCCGACCTTAACCTGCGAGTCACCAACACGACGACGAGCAGCTCCAGC 1397
QY	21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40
Db	1398 TGTGACGAGGCGCTCTTCGCTCCGGCCCGCCACTCGTCCCGCTGGCGAGGACGGT 1457
QY	41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60
Db	1458 CGTGTGAGCCCTATGTGGACTTGTCTGAGTTTTTCCGCTCTGGAGGTGGACCATGGC 1517
QY	61 GluGlnSerValValThrAlaPro 68
Db	1518 GAGCAGCGTGGTGACGACCG 1541
RESULT	9
LOCUS	HSU49928
DEFINITION	Homo sapiens TAK1 binding protein (TAB1) mRNA, complete cds.
ACCESSION	U49928
VERSION	U49928.1 GI:1401125
KEYWORDS	
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	Shibuya,H., Yamaguchi,K., Shirakabe,K., Tonegawa,A., Gotoh,Y.,
JOURNAL	Ueno,N., Irie,K., Nishida,E. and Matsumoto,K.
MEDLINE	TAB1: an activator of the TAK1 MAPKKK in TGF-beta signal
PUBMED	transduction
REFERENCE	Science 272 (5265), 1179-1182 (1996)
AUTHORS	Shibuya,H.
TITLE	Direct Submission
JOURNAL	Submitted (25-FEB-1996) Hiroshi Shibuya, Faculty of Pharmaceutical
FEATURES	Sciences, Hokkaido University, Nishi 6-chome, Kita 12, Kita-Ku,
source	Sapporo, Hokkaido 060, Japan
	Location/Qualifiers
	1..3096



```
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3096
/genre="TAB1"
21..1535
/feature="TAB1"
/note="activator for TAK1"
/codon_start=1
/product="TAK1 binding protein"
/protein_id="PAC12660.1"
/db_xref="GI:1401126"
/translation="MAQRSLQSQEQSWTDPLCHLSGSGSASNSRSADKGT
ESHPEDSMLEFRENKCFLYGVNGYDGNRTYNAQRLSELLQLEHAEAY
RVLLOAPDVVERFLESIDDALEKASLOSOLPEGVPHOHLPPOYKILRLTLER
EISGGAIVAVVILNKLIVYANTGRTALCKSTYGVQVNTDFTTEDELFPL
SOLGDAGKIKOVGILTCGSHSTRIDGVKYGYNDIDLSAKSPKPIIAPERTHGO
PLDGVTFVLMSEGLKALERAHGSGQANOEIAMAIDEPKQTSLDAYVQAVVDKY
KRHSDFASGGERARCPRHEDWTLVRNFGPLGEMSOPTSPAPAPAGRVPSV
PYSSAQSTSKSTVTLISLMPGQMVNGAHASLTIDEATPTLTNOSPVLTIQNTHT
QSSSSSDGGLFRSRPHSLPPEGDRVEPYDFAEFYALMSVDHGSVVTAP"
BASE COUNT      642 a      936 c      952 g      565 t
ORIGIN
Alignment Scores:      4.14e-30      Length:      3096
Pident:      359.00      Matches:      68
Score:      100.00%      Conservative:      0
Percent Similarity:      100.00%      Mismatches:      0
Best Local Similarity:      100.00%      Indels:      0
Query Match:      9      Gaps:      0
DB:
US-09-830-144-4_COPY_437_504 (1-68) x HSU49928 (1-3096)
QY      1      GlnSerProThleuThleuGlnSerThraAnthrhsthrGlnSerSerSerSer 20
Db      1329      CAAAGCCGACCTTAACTGACGTCACCAACGACGACGACGACGACCTCCAGC 1388
QY      21      SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGly 40
Db      1389      TCTGACGAGGAGGCTCTCCGCTCCGCGCCGCCACCTCGCTCCGCGCTGCGAGGACGGT 1448
QY      41      ArgValGluProTyrValAlaPheAlaGluPheTyrArgLeuTyrPsrValaPheHisGly 60
Db      1449      CGGTTGAGCCCTATGTGACATTTGCTGATTAACGCCCTCGAGACGTGACCATGGC 1508
QY      61      GluGlnSerValValThrAlaPro 68
Db      1509      GAGCAGAGGGGTGTGACACGACCG 1532
RESULT 10
HS407F17      69660 bp      DNA      linear      PRI 12-DEC-1999
LOCUS      HS407F17
DEFINITION      Human DNA sequence from clone R3J-407F17 on chromosome 22 contains
the gene for TAB1 (TAK1 binding protein 1), ESTs, STSs, GSSs and
two putative CpG islands, complete sequence.
ACCESSION      Z83845
VERSION      Z83845.14
KEYWORDS      GI:5441636
SOURCE      HTG; CpG island; TAB1; TAK1.
ORGANISM      Homo sapiens
Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
Makariota; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 69660)
AUTHORS      Sanger Centre, Hinxton, Cambridgeshire,
TITLE      Direct Submission
JOURNAL      Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT      Requested by Sanger, ac.uk
On Jul 10, 1999 this sequence version replaced gi:5419637.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
```

```
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
R3J-407F17 is from the library R3C1-3 constructed at the Roswell Park Cancer
Institute by the group of Pieter de Jong. For further details see
http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
R3J-407F17. It may be shorter because we only sequence overlapping
sections once, or longer because we arrange for a small overlap
between neighbouring submissions.
The true left end of clone R3J-407F17 is at 1 in this sequence. The
true left end of clone R3J-1104E15 is at 69561 in this sequence.
The true right end of clone R3J-33H23 is at 1123 in this
sequence. The start of this sequence overlaps with sequence
AL022326.
FEATURES
source      location/Qualifiers
1..69660
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="R3J-407F17"
/clone_1lb="R3C1-3"
30..216
/note="AluJb repeat: matches 111..296 of consensus"
31..412
/note="match: STS: Em:HS407F17S; match: STS: Em:AL021668"
346..437
/note="MER45B repeat: matches 1..74 of consensus"
438..753
/note="AluX repeat: matches 1..308 of consensus"
754..790
/note="MER45B repeat: matches 74..114 of consensus"
879..1189
/note="AluSg repeat: matches 1..313 of consensus"
1198..1507
/note="AluSg repeat: matches 1..298 of consensus"
1855..2147
/note="AluSx repeat: matches 1..293 of consensus"
2367..2845
/note="match: GSS: Em:AQ877061"
2485..2657
/note="AluSg/x repeat: matches 135..308 of consensus"
3146..3341
/note="MER45B repeat: matches 662..1039 of consensus"
3504..3817
/note="AluY repeat: matches 1..311 of consensus"
3820..3836
/note="WIR repeat: matches 193..207 of consensus"
3837..4187
/note="AluSx repeat: matches 1..312 of consensus"
4188..4356
/note="WIR repeat: matches 8..193 of consensus"
4613..4797
/note="AluSg/x repeat: matches 126..306 of consensus"
4864..5019
/note="L2 repeat: matches 1599..1763 of consensus"
5026..5206
/note="L2 repeat: matches 2241..2420 of consensus"
```

```
repeat_region 5293. .5594
/note="AluYb8 repeat: matches 1. .308 of consensus"
5597. .5890
/note="AluJb repeat: matches 12. .303 of consensus"
5891. .6057
/note="AluSp repeat: matches 126. .300 of consensus"
6068. .6236
/note="L2 repeat: matches 2544. .2750 of consensus"
6539. .7149
/note="L2 repeat: matches 1815. .2541 of consensus"
7371. .7432
/note="MIR repeat: matches 158. .219 of consensus"
7439. .7516
/note="AluJb repeat: matches 1. .60 of consensus"
7517. .7691
/note="AluSc repeat: matches 132. .306 of consensus"
7692. .7986
/note="AluSg repeat: matches 1. .294 of consensus"
7987. .8213
/note="AluJb repeat: matches 53. .281 of consensus"
8297. .8506
/note="AluSp repeat: matches 2. .212 of consensus"
8691. .8994
/note="AluX repeat: matches 1. .301 of consensus"
9055. .9367
/note="AluSg repeat: matches 1. .310 of consensus"
9532. .9655
/note="AluSg/x repeat: matches 69. .192 of consensus"
9656. .9804
/note="AluJb repeat: matches 149. .297 of consensus"
9873. .10415
/note="CpG island"
/evidence=not_experimental
10049. .42131
/gene="dJ407F17.1"
mRNA
join(10049. .10084,25255. .25391,25749. .25902,27040. .27126,
27960. .28098,28981. .29094,29768. .29879,32076. .32220,
36952. .37174,38270. .38432,40264. .42131)
/gene="dJ407F17.1"
/product="dJ407F17.1 (TAB1 (TAK1 binding protein 1))"
/note="match: cDNA: Em:U92031 Em:U49928; match: ESTs:
Em:U17295 Em:AA726609 Em:AA019355 Em:AA168567 Em:AA1654435
Em:AA789149 Em:AI014848 Em:AA430502 Em:AA745822
Em:AA411895 Em:AA411894 Em:AA326751 Em:AA430621 Em:H59337
Em:AA430462 Em:AA806521 Em:R40486 Em:AA019356 Em:AA157525
Em:AA028980 Em:T34667 Em:R83147 Em:A1269211 Em:AA806517
Em:AA147685 Em:AA458466 Em:Z43346 Em:A1656463 Em:W95352
Em:AA828095 Em:R60090 Em:AA280338 Em:AA679208 Em:AA442978
Em:F08391 Em:AI085576 Em:W95394 Em:RI3035 Em:H48607
Em:R60028 Em:F02762"
/evidence=not_experimental
CDS
join(10052. .10084,25255. .25391,25749. .25902,27040. .27126,
27960. .28098,28981. .29094,29768. .29879,32076. .32220,
36952. .37174,38270. .38432,40264. .40471)
/gene="dJ407F17.1"
/note="match: proteins: Tr:Q15750 Tr:O73614"
/codon_start=1
/evidence=not_experimental
/product="dJ407F17.1 (TAB1 (TAK1 binding protein 1))"
/protein_id="CA855304.1"
/db_xref="GI:5834565"
/translation="MAAQRRLSQEQPQSDTDLPLCHLSGVGSASNRYSADGKGT
RHPVLPDLSLKRNSNCFYGVNGYDGNRVNFAVRSLASALLGQLNASHADVD
EVLQAQDVVERSPLEISIDDALEAKASQLPQGVPOHLPQYOKLTLEKLTLE
ETSGGMAVAVLANKLYVANVTGNALLCKSTVDGLQVLTQNVDPHTENDELFR
SOLGLDAGIKVGVGIIICQESTRIGYKVKYGYTIDLLSAKSPFIIAREIHQAQ
PLDGVTFGLVLMSEGLYKALEARHPGQANQETAMIDTEFAKQSLDAVAQVDRV
KRHSDFASGTERARFCPRHEDMTLLVNRFPGLGEMSOFTSPAPAGGVRVPSV
PVSAAQSTKSVTLVLYMPQGGQVNGASHTLDEATPTLNQSTLTQSTNTH
QSSSSSDGLFRPSRAHSLPFGDGRVPEYVFAEYRLWSVDHGEQSVTAP"
10423. .10665
/note="MIR repeat: matches 3. .261 of consensus"
10894. .11003
repeat_region
```

## AUTHORS

Strausberg, R.  
Direct Submission  
Submitted (04-Apr-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

## REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

## COMMENT

Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
Sequencing Center by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Huijck, S.W., Hale, S.M.,  
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://lml.lnl.gov>  
Series: IRAX Plate: 45 Row: K Column: 11  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis, Genomescan gene prediction.

## FEATURES

## SOURCE

Location/Qualifiers  
1. 2944  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="FVB/N"  
/clone="IMAGE:5042323"  
/issue\_type="Kidney, normal, 5 month old male mouse."  
/clone\_1fb="NCI CGAP\_Kid14"  
/1ab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
<1. 1503  
/codon\_start=1  
/product="Similar to mitogen-activated protein kinase  
kinase kinase 7 interacting protein 1"  
/protein\_id="AAH27054.1"  
/db\_xref="GI:20072134"  
/translation="AQRSLQSQSPQSWTDDPLCLTSGVGSASNSYSADKGTES  
HPEDNMLKFRSENNCPFLGVNGVDGNRVNPFVAQPLSMLLQCLNTEHLEAVR  
VLLQADVVERFESIDDALEAKASLOSOLPEVPOHLPQYOKILRLKALREI  
SGAMAVAVILNSKLYVANYGTRALLCKSTVDGQVOTQIMNDHTTEDELPLSQ  
LGLDAKIKOMGVICGOSPRIGDYKYKQYDIDILSLAKSKPIIARPTHTGQPL  
DQVTFPLVMSBGLKALBAHAGPQANQELIANDTEPAKQSLDAVAQAVDYKX  
HSDTPASGSEKAPCFRHEHMTLLVKNFGPIGEMSOPTPTAPGGRVYPVSVYS  
AOSTKSTVLSLVMPISQGVNNGSHASTLDEATPTLTIQSPITLTIQSTNTHQSSS  
SSSDGLFSPRPAHSLPGEDEGRVEPYFAEFYRLMSVDHGSQVMTAP"

## CDS

BASE COUNT 676 a 835 c 863 g 570 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 8,426-30 Length: 2944  
Score: 356.00 Matches: 67  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.53% Mismatches: 0  
Query Match: 99.16% Indels: 0  
Gaps: 0  
DB: 10

US-09-630-144-4\_COPY\_437\_504 (1-68) x BC027054 (1-2944)

OY 1 Glnserpochrleuthrleuglnsertthraanthrhthglnserserserser  
DB 1297 CAGAGCCCATCTGACCTCGACGTCCACCAACGACACCAAGAGAGCTCCAGC 1356  
OY 21 Seraspjlgyljeuphearsersatpcoalaahserserleupropiglyglaaspjly 40  
DB 1357 TCTGACGGGGGCTCTTCGCTCCAGACCGGCTCACTCACTTCCACCCGAGAGAGATGGC 1416

OY 41 ArgValGlpicrtyValAspPheAlaGluPheTyArgLeuTrpserValAspHisctly 60  
DB 1417 CGGTGTGAGCCCTATGTGACTTGTCTGAGTTCTACCGACTCTGGAGCGGTGACACGGC 1476  
OY 61 GluGlnserValValThrAlaPro 68  
DB 1477 GACCAACGCTGATGACGCGCACT 1500

## RESULT 12

## AC127924/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

1 (bases 1 to 141498)  
Muzny, D.M., Adams, C., Adio-Odojola, B., Ali-osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Ayala, M., Banks, T.,  
Barbata, V., Bencon, J., Bimberg, K., Blankenburg, R., Bonnin, D.,  
Bouck, U., Bowe, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dahorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Dounswalte, K.U., Drepper, H., Dugan-Kocha, S., Durbin, K.J.,  
Earmhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frenzt, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Haylak, P., Hayes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,  
Homs, F., Howard, S., Huber, J., Huijck, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S.,  
Karlsson, E., Kelly, S., Khan, U., King, L., Kovach, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,  
Lozdo, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabhat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokemwo, S., Ogdu, M., Okunolu, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
Scherer, S., Scott, G., Shen, H., Shooshbari, N., Sisson, I.,  
Sodergren, E., Sonajke, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svaltek, A., Tabot, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, U., Vera, V., Villalob, D., Vinson, R., Wang, S.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, C.,  
Williams, G., Williamson, A., Wleczky, R., Woodson, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G., and Gibbs, R.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## COMMENT

Direct Submission  
Submitted (19-JUN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
----- Genome Center  
Center: Baylor College of Medicine

```

Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center Project name: KAAS
Center clone name: CH230-23318
----- Summary Statistics -----
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 86446 bases at least Q40
Consensus quality: 92843 bases at least Q30
Consensus quality: 96514 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1030: contig of 1030 bp in length
* 1031: 1030: gap of unknown length
* 1131: 2262: contig of 1132 bp in length
* 2263: 2262: gap of unknown length
* 2363: 3440: contig of 1078 bp in length
* 3441: 3540: gap of unknown length
* 3541: 5167: contig of 1627 bp in length
* 5168: 5267: gap of unknown length
* 5268: 6580: contig of 1313 bp in length
* 6581: 6680: gap of unknown length
* 6681: 7883: contig of 1203 bp in length
* 7884: 7983: gap of unknown length
* 7984: 9021: contig of 1038 bp in length
* 9022: 9121: gap of unknown length
* 9122: 10825: contig of 1704 bp in length
* 10826: 10925: gap of unknown length
* 10926: 12083: contig of 1158 bp in length
* 12084: 12183: gap of unknown length
* 12184: 13310: contig of 1127 bp in length
* 13311: 13410: gap of unknown length
* 13411: 14464: contig of 1054 bp in length
* 14465: 14564: gap of unknown length
* 14565: 16304: contig of 1740 bp in length
* 16305: 16404: gap of unknown length
* 16404: 18339: contig of 1935 bp in length
* 18340: 18439: gap of unknown length
* 18440: 19493: contig of 1054 bp in length
* 19494: 19593: gap of unknown length
* 19594: 21298: contig of 1705 bp in length
* 21299: 21398: gap of unknown length
* 21399: 22537: contig of 1139 bp in length
* 22538: 22637: gap of unknown length
* 22638: 24129: contig of 1492 bp in length
* 24130: 24229: gap of unknown length
* 24230: 25277: contig of 1048 bp in length
* 25278: 25377: gap of unknown length
* 25378: 27078: contig of 1701 bp in length
* 27079: 27178: gap of unknown length
* 27179: 29508: contig of 2330 bp in length
* 29509: 29608: gap of unknown length
* 29609: 31420: contig of 1812 bp in length
* 31421: 31520: gap of unknown length
* 31521: 33234: contig of 1714 bp in length
* 33235: 33334: gap of unknown length
* 33335: 35110: contig of 1776 bp in length
* 35111: 35210: gap of unknown length
* 35211: 36675: contig of 1465 bp in length
* 36676: 36775: gap of unknown length
* 36776: 37894: contig of 1119 bp in length
* 37895: 37994: gap of unknown length
*
* 37995: 39970: contig of 1976 bp in length
* 39971: 40070: gap of unknown length
* 40071: 41623: contig of 1553 bp in length
* 41624: 41723: gap of unknown length
* 41724: 43157: contig of 1434 bp in length
* 43158: 43257: gap of unknown length
* 43258: 45044: contig of 1787 bp in length
* 45045: 45144: gap of unknown length
* 45145: 46604: contig of 1460 bp in length
* 46605: 46704: gap of unknown length
* 46705: 47898: contig of 1194 bp in length
* 47899: 47998: gap of unknown length
* 47999: 49208: contig of 1210 bp in length
* 49309: 49308: gap of unknown length
* 49309: 50615: contig of 1307 bp in length
* 50616: 50715: gap of unknown length
* 50716: 52499: contig of 1784 bp in length
* 52500: 52599: gap of unknown length
* 52500: 54003: contig of 1404 bp in length
* 54004: 54103: gap of unknown length
* 54104: 55661: contig of 1558 bp in length
* 55662: 55761: gap of unknown length
* 55762: 58110: contig of 2349 bp in length
* 58111: 58210: gap of unknown length
* 58211: 61322: contig of 3112 bp in length
* 61323: 61422: gap of unknown length
* 61423: 62808: contig of 1386 bp in length
* 62809: 63808: gap of unknown length
* 63809: 65459: contig of 2551 bp in length
* 65460: 65559: gap of unknown length
* 65560: 66882: contig of 1323 bp in length
* 66883: 66982: gap of unknown length
* 66983: 68456: contig of 1474 bp in length
* 68457: 68556: gap of unknown length
* 68557: 70789: contig of 2233 bp in length
* 70790: 70889: gap of unknown length
* 70890: 73103: contig of 2214 bp in length
* 73104: 73203: gap of unknown length
* 73204: 74721: contig of 1518 bp in length
* 74722: 74821: gap of unknown length
* 74822: 77416: contig of 2595 bp in length
* 77417: 77516: gap of unknown length
* 77517: 78949: contig of 1433 bp in length
* 78950: 79049: gap of unknown length
* 79050: 81501: contig of 2452 bp in length
* 81502: 81601: gap of unknown length
* 81602: 83009: contig of 1408 bp in length
* 83010: 83109: gap of unknown length
* 83110: 84773: contig of 1664 bp in length
* 84774: 84873: gap of unknown length
* 84874: 86891: contig of 2018 bp in length
* 86892: 86991: gap of unknown length
* 86992: 88782: contig of 1791 bp in length
* 88783: 88882: gap of unknown length
* 88883: 90589: contig of 1707 bp in length
* 90590: 90689: gap of unknown length
* 90690: 92680: contig of 1991 bp in length
* 92681: 92780: gap of unknown length
* 92781: 94508: contig of 1728 bp in length
* 94509: 94608: gap of unknown length
* 94609: 97262: contig of 2654 bp in length
* 97263: 97362: gap of unknown length
* 97363: 100119: contig of 2757 bp in length
* 100120: 100219: gap of unknown length

```

## Alignment Scores:

```

Pred. No.: 4.52e-28      Length: 14198
Score: 356.00           Matches: 67
Percent Similarity: 100.00%      Conservative: 1
Best Local Similarity: 98.53%     Mismatches: 0
Query Match: 99.16%             Indels: 0
DB: 2                       Gaps: 0

```

```

US-09-830-144-4_COPY_437_504 (1-68) x AC127924 (1-141498)
Oy      1  GlnserProThrluInserThrAsnThrHisThGlnSerSerSerSer 20
Db 125460 CAGAGCCCCCTGACCTGACGTCTACCAACACCAACCAAGGACGACCTCCAGC 125401
Oy      21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGlyAspGly 40
Db 125400 TCGGACGGGGGCGCTCTCCGCTCCAGACCGGCTCATCTTCCACTTCCAGAGAGATGCG 125341
Oy      41 ArgValGluProGlyTrpValaPheAlaGluPheTrpArgLeuTrpSerValaAspHisGly 60
Db 125340 AGGGTCGAGCCCTATGTGGAGCTTTGCTGAGAGTTTACGACTCGAGGCGTGCATCATGCG 125281
Oy      61 GluInserValaValThrAlaPro 68
Db 125280 GAGCGAGGTGATGATGACGCGACCT 125257

RESULT 13
AC127784 176665 bp DNA linear HTG 19-JUN-2002
LOCUS Rattus norvegicus clone CH230-131B6, *** SEQUENCING IN PROGRESS
DEFINITION *** 59 unordered pieces.
AC127784
VERSION AC127784.1 GI:21908163
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
REFERENCE 1 (bases 1 to 176665)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,
          Albrooks,S.B., Amaralunge,H.C., Aye,J.R., Ayele,M., Banks,F.,
          Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
          Butch,J., Bowle,S., Brieva,M., Brown,M., Bryant,N.P.,
          Butay,C., Butch,P., Butrett,C., Butrell,K.L., Byrd,N.C.,
          Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
          Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
          Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
          Delaney,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
          Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
          Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
          Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
          Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
          Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
          Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
          Harrie,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
          Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
          Homsi,F., Howard,S., Huber,U., Huiyk,S., Hume,J., Jackson,L.E.,
          Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
          Katsionis,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
          Kretovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
          Li,J., Li,Z., Licharge,O., Lien,C., Liu,J., Liu,W., Louisged,H.,
          Lorado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
          Mathewarati,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
          Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzke,M.,
          Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
          Moser,M., Neal,D., Newton,J., Newton,S.N., Nguyen,A., Nguyen,N.,
          Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okumura,G.,
          Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
          Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
          Rivers,M., Rojas,A., Rojurokan,I., Rolfe,M., Ruiz,S., Savary,G.,
          Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
          Soederren,E., Sonatke,T., Sparks,A., Stanley,H., Stone,H.,
          Sutton,A., Svatek,A., Tabor,P., Tamezisa,A., Tamezisa,K., Tang,H.,
          Tatum,J., Taylor,C., Taylor,T., Teleford,B., Thomas,N., Thomas,S.,
          Usmani,K., Vasquez,L., Vera,Y., Villalón,D., Vinson,R., Wang,Q.,
          Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
          Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
          Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
          Weinstein,G. and Gibbs,R.
          Direct Submission

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176665)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          -----
          Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          -----
          Project Information
          Center project name: GZWK
          Center clone name: CH230-131B6
          -----
          Summary Statistics
          Sequencing vector: Plasmid;
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 120196 bases at least Q40
          Consensus quality: 127192 bases at least Q30
          Consensus quality: 131862 bases at least Q20
          -----
          * NOTE: Estimated insert size may differ from sequence length
          * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 59 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          1
          1032: contig of 1032 bp in length
          1033
          1132: gap of unknown length
          1133
          2439: contig of 1307 bp in length
          2440
          2539: gap of unknown length
          2540
          3610: contig of 1071 bp in length
          3611
          3710: gap of unknown length
          3711
          4955: contig of 1245 bp in length
          4956
          5055: gap of unknown length
          5056
          7048: contig of 1993 bp in length
          7049
          7148: gap of unknown length
          7149
          8902: contig of 1754 bp in length
          8903
          9002: gap of unknown length
          9003
          10567: contig of 1565 bp in length
          10568
          10667: gap of unknown length
          10668
          12258: contig of 1591 bp in length
          12259
          12358: gap of unknown length
          12359
          14216: contig of 1858 bp in length
          14217
          14316: gap of unknown length
          14317
          15760: contig of 1444 bp in length
          15761
          15860: gap of unknown length
          15861
          16935: contig of 1075 bp in length
          16936
          17035: gap of unknown length
          17036
          18984: contig of 1949 bp in length
          18985
          18984: gap of unknown length
          18985
          20295: contig of 1211 bp in length
          20296
          20395: gap of unknown length
          20396
          21672: contig of 1277 bp in length
          21673
          21772: gap of unknown length
          21773
          23061: contig of 1289 bp in length
          23062
          23161: gap of unknown length
          23162
          24596: contig of 1435 bp in length
          24597
          24596: gap of unknown length
          24597
          26381: contig of 1685 bp in length
          26382
          26481: gap of unknown length
          26482
          29059: contig of 2578 bp in length
          29060
          29159: gap of unknown length
          29160
          30322: contig of 1163 bp in length
          30323
          30422: gap of unknown length
          30423
          32594: contig of 2172 bp in length
          32595
          32694: gap of unknown length

```

TITLE

* 32695	34702:	contig of 2008 bp in length
* 34703	34802:	gap of unknown length
* 34803	37016:	contig of 2214 bp in length
* 37017	37116:	gap of unknown length
* 38121	contig of 1205 bp in length	
* 38421	gap of unknown length	
* 38422	contig of 2428 bp in length	
* 40849	contig of unknown length	
* 40850	gap of unknown length	
* 43133	contig of 2184 bp in length	
* 43234	gap of unknown length	
* 43386	contig of 2153 bp in length	
* 45387	gap of unknown length	
* 45877	contig of 2890 bp in length	
* 48376	gap of unknown length	
* 48377	contig of 2931 bp in length	
* 51407	gap of unknown length	
* 51507	gap of unknown length	
* 53492	contig of 1985 bp in length	
* 53493	gap of unknown length	
* 53593	contig of 2185 bp in length	
* 55777	gap of unknown length	
* 55778	contig of 2776 bp in length	
* 58653	gap of unknown length	
* 58753	gap of unknown length	
* 58754	contig of 2506 bp in length	
* 61259	gap of unknown length	
* 61360	contig of 1834 bp in length	
* 63194	gap of unknown length	
* 63294	contig of 3188 bp in length	
* 66481	gap of unknown length	
* 66581	gap of unknown length	
* 68046	contig of 1465 bp in length	
* 68146	gap of unknown length	
* 72431	contig of 4285 bp in length	
* 72531	gap of unknown length	
* 72532	contig of 2313 bp in length	
* 74844	gap of unknown length	
* 74845	contig of 3007 bp in length	
* 77951	gap of unknown length	
* 77952	contig of 1613 bp in length	
* 81664	gap of unknown length	
* 81665	gap of unknown length	
* 81666	contig of 2924 bp in length	
* 84688	gap of unknown length	
* 84788	gap of unknown length	
* 87987	contig of 3199 bp in length	
* 87988	gap of unknown length	
* 91325	contig of 3238 bp in length	
* 91326	gap of unknown length	
* 94907	contig of 3482 bp in length	
* 95007	gap of unknown length	
* 95008	contig of 3182 bp in length	
* 95009	gap of unknown length	
* 98289	gap of unknown length	
* 98290	contig of 4001 bp in length	
* 102291	gap of unknown length	
* 102390	contig of 2829 bp in length	
* 105219	gap of unknown length	
* 105220	contig of 3214 bp in length	
* 108533	gap of unknown length	
* 108634	contig of 3505 bp in length	
* 112139	gap of unknown length	
* 112139	contig of 3901 bp in length	
* 116140	gap of unknown length	
* 116239	gap of unknown length	
* 116240	contig of 4988 bp in length	
* 121227	gap of unknown length	
* 121228	contig of 4119 bp in length	
* 121328	gap of unknown length	
* 125446	contig of 4423 bp in length	
* 125447	gap of unknown length	
* 125969	contig of 3423 bp in length	
* 129069	gap of unknown length	
* 129070	contig of 4713 bp in length	
* 133782	gap of unknown length	
* 133783	contig of 7183 bp in length	
* 141065	gap of unknown length	
* 141165	contig of 6029 bp in length	
* 141166	gap of unknown length	
* 147194	contig of 7101 bp in length	
* 147195	gap of unknown length	
* 154395	contig of 1701 bp in length	
* 154396	gap of unknown length	
* 154496	contig of 6306 bp in length	

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

5.68e-28

356.00

100.00%

98.53%

99.16%

2

Length:

Matches:

Conservative:

Mismatches:

Indels:

Gaps:

176665

67

1

0

0

US-09-830-144-4\_COPY\_437\_504 (1-68) x AC127784 (1-176665)

OY

1

GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSer

20

Db

163208

CAGAGCCCCACTCTGACCTCGAGTCTACCAACACCCACACAGGACGACCTCCAGC

163267

OY

21

SerAspGlyClyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly

40

Db

163268

TCGAGCGGGGGCTCTCCGCTCCAGACCGGCTCACCTCACCTCGAGAGGATGGC

163327

OY

41

ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly

60

Db

163328

AGGTCGACCCCTATGTGACTTTCTGAGTTTACCGACTCTGGAGCGTGGACCATGGC

163387

OY

61

GluGlnSerValValThrAlaPro

68

Db

163388

GAGCAGAGTGTGATGACGGCACCT

163411

RESULT 14

XLU92031

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

BASE COUNT

ORIGIN

160802

160901:

gap of unknown length

160801:

contig of 6306 bp in length



Search completed: December 10, 2002, 06:49:53  
Job time : 3161 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: December 10, 2002, 04:05:41 ; Search time 319 Seconds  
(without alignments)  
480.050 Million cell updates/sec

Title: US-09-830-144-4\_COPY\_437\_504

Perfect score: 359 1 QSFPTLTLGSTNTHGTSSSSS.....AEFYRLMSVDHGOSVVTAP 68

Sequence: 1 QSFPTLTLGSTNTHGTSSSSS.....AEFYRLMSVDHGOSVVTAP 68

Scoring table: BLOSUM62  
Xgapod 10.0 , Xgapext 0.5  
Ygapod 10.0 , Ygapext 0.5  
Fgapod 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODBL=frame+ p2n model -DEV=xip  
-O=/cgnr\_1/USPTO.spool/US09830144/runat\_04122002\_141842\_4964/app\_query.fasta\_1.263  
-DB=Geneseq\_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09830144 @CGN 1.1 79 @runat 04122002 141842 4964 -NCPU=6 -ICPU=3  
-NO XLPY -NO MMAP -LARGJQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOD=10 -XGAPEXT=0.5 -FGAPOD=6 -FGAPEXT=7  
-YGAPOD=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseg/geneseg-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	359	100.0	1515	21	AA248861 Human TAB1 coding
2	359	100.0	1560	18	AA291175 Human TAB1 (TAK1) b
3	359	100.0	1560	18	AA291178 Human TAB1 (TAK1) b
4	359	100.0	1560	20	AA56278 Human TAB1 encodin
5	359	100.0	1560	21	AAA39106 Human TAB-1 nucleo
6	359	100.0	1568	20	AA562310 Human TAB1 encodin
7	359	100.0	1569	20	AA56282 Human TAB1-FLAG en
8	359	100.0	1687	22	ABA20494 Human nervous syst
9	359	100.0	1687	22	AA136984 Human prostatic can
10	352	98.1	696	21	AA15895 Human prostate relat
11	76	21.2	519	24	ABN61139 Drosophila melanog
12	75	20.9	14091	23	ABU11586 Drosophila melanog
13	74.5	20.8	990	23	AA580813 DNA encoding novel
14	71.5	19.8	1108	14	AA049396 Class II AP endonu
15	71	19.8	996	24	ABK78528 Bacillus clausii g
16	70	19.5	3489	23	ABU28043 Drosophila melanog
17	70	19.5	19674	23	ABU28042 Drosophila melanog
18	69.5	19.4	1492	22	AAH78731 Human HT4S64 seri
19	69.5	19.4	3820	23	ABU12378 Drosophila melanog
20	69.5	19.4	4877	23	ABU13708 Human reproductive
21	69.5	19.4	9507	22	AAU07097 DNA encoding novel
22	69	19.2	6743	23	AA565172 DNA encoding novel
23	69	19.2	6743	23	AA588306 DNA encoding novel
24	69	19.2	12409	22	AA530238 Human immune/haema
25	69	19.2	23580	22	AAK66230 Human phosphoinosi
26	69	19.2	23580	22	AAK83578 Human phosphoinosi
27	68.5	19.1	3068	22	AA514724 Human phosphoinosi
28	68.5	19.1	12094	23	AA595353 Protonibacterium
29	68.5	19.1	16951	22	AAH48620 Human fescin DNA f
30	68.5	19.1	16951	22	AAH48622 Human fescin DNA f
31	68.5	19.1	16951	22	AAH48622 Human fescin DNA f
32	68	18.9	273	22	ABA70749 Human foetal liver
33	68	18.9	273	22	AAK19001 Human brain expres
34	68	18.9	273	22	AAK4951 Human bone marrow
35	68	18.9	273	22	AA125076 Probe #15009 for g
36	68	18.9	273	22	AA150921 Human genome-deriv
37	68	18.9	273	24	AB519195 Human genomic DNA
38	68	18.9	8918	22	AA526708 Human genomic DNA
39	68	18.9	8918	22	AA526709 Human genomic DNA
40	67.5	18.8	582	21	AA044783 zea mays DNA fragm
41	67.5	18.8	2217	23	AA580664 DNA encoding novel
42	67.5	18.8	2377	22	ABA82579 Atrophin-1 interac
43	67.5	18.8	2559	20	AA029236 Human E3 ubiquitin
44	67.5	18.8	2559	20	AA029236 Human E3 ubiquitin
45	67.5	18.8	5372	20	AA029235 Human E3 ubiquitin

## ALIGNMENTS

RESULT 1  
AA248861  
ID AA248861 standard; CDNA: 1515 bp.

AA248861;  
24-MAR-2000 (first entry)

Human TAB1 coding sequence.

Human, TAB1, XIAP, X-linked inhibitor of apoptosis protein; TGF-beta;  
transforming growth factor-beta activated kinase 1; monocyte migration;  
TAK1 binding protein 1; extracellular matrix protein production;  
cell growth inhibitor; beta-amyloid protein deposition;  
immunopression; Transforming growth factor-beta; ds.

Homo sapiens.

PN JPL1326328-A.  
 XX 26-NOV-1999.  
 XX 13-MAY-1998; 98JP-0130378.  
 XX 13-MAY-1998; 98JP-0130378.  
 XX (MATS/) MATSUMOTO K.  
 XX WPI; 2000-078337/07.  
 DR P-PSDB; AAY59450.  
 XX  
 PT Screening a substance which inhibits combination of the X-linked  
 PT inhibitor of apoptosis protein -  
 XX  
 PS Claim 2; Page 25-26; 43pp; Japanese.  
 XX  
 CC This sequence encodes the human TAB1 protein.  
 CC The invention relates to a method for screening a substance inhibiting  
 CC the formation of a complex between XIAP and TAB1, in which X-linked  
 CC inhibitor of apoptosis protein (XIAP), transforming growth factor-beta  
 CC activated kinase 1 (TAK1) binding protein 1 (TAB1) and a substance to be  
 CC tested are contacted with each other and then the presence or formation  
 CC of a complex between XIAP and TAB1 is detected. The substance can be used  
 CC as a drug for extracellular matrix protein production enhancement, cell  
 CC growth inhibition, monocyte migration, physiologically active substance  
 CC induction, immunosuppression, and beta-amyloid protein deposition. A  
 CC substance inhibiting the formation of a complex between TAB1 and XIAP as  
 CC well as between XIAP and TGF-beta (transforming growth factor-beta) type  
 CC I and/or type II receptor is useful as a drug.  
 XX  
 SQ Sequence 1515 BP; 323 A; 457 C; 463 G; 272 T; 0 other;

Alignment Scores:  
 Pred. No.: 2,336-35 Length: 1515  
 Score: 359.00 Matches: 68  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x AA248861 (1-1515)

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
 |||||  
 Db 1309 CAAGCCCGACCTTAACCTTCAGTCCACCAACAGCGACGACGAGCAGCTCCAGC 1368  
 |||||

Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40  
 |||||  
 Db 1369 TCTGACGGAGGCGCTCTTCGCTCCCGGCCGCCCACTCGCTCCCGCTGGCGAGGACGGT 1428  
 |||||

Qy 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTyrSerValAspHisGly 60  
 |||||  
 Db 1429 CGTGTTGAGCCCTATGTGACCTTCTGAGTTTACCGCCTCTCGAGCGTGGACCATGGC 1488  
 |||||

Qy 61 GluGlnSerValValThrAlaPro 68  
 |||||  
 Db 1489 GAGCAGAGCGTGTGACGACCCG 1512

RESULT 2  
 AAT91175  
 ID AAT91175 standard; cDNA; 1560 BP.  
 XX  
 AC AAT91175;  
 XX  
 DT 14-APR-1998 (first entry)  
 XX  
 DE Human TAB1 (TAK1 binding protein) cDNA.  
 XX  
 KW TAB1; TAK1 binding protein; transforming growth factor-beta;  
 KW signal transduction; human; ds.  
 XX

OS Homo sapiens.  
 XX Key Location/Qualifiers  
 XX CDS 30..1544  
 FT /\*tag= a  
 FT variation 185  
 FT /\*tag= b  
 FT /note= "another clone has adenine at position 185,  
 FT with codon AGC (Ser) altered to AGA (Arg)"  
 XX  
 XX EP803571-A2.  
 XX PD 29-OCT-1997.  
 XX  
 PF 24-APR-1997; 97EP-0302808.  
 XX  
 PR 20-NOV-1996; 96US-0752891.  
 PR 24-APR-1996; 96JP-0126282.  
 PR 28-OCT-1996; 96JP-0300856.  
 XX  
 XX (UENO/) UENO N.  
 XX  
 XX Matsumoto K, Nishida E;  
 XX WPI; 1997-515318/48.  
 DR P-PSDB; AAW26706.  
 XX  
 CC DNA encoding TAK1 binding protein TAB1 - member of transforming  
 CC growth factor beta receptor signal production pathway, which  
 CC activates TAK-1 kinase activity upon binding  
 XX  
 PS Claim 1; Page 17-19; 30pp; English.  
 XX  
 CC This cDNA clone codes for human TAB1 (see AAW26706), a novel member  
 CC of the transforming growth factor-beta receptor signal transduction  
 CC pathway, which activates TAK-1 kinase activity upon binding. To  
 CC obtain the full-length TAB1 coding sequence, a human kidney cDNA  
 CC library was screened using as a probe a partial TAB1 cDNA obtained  
 CC from a yeast two-hybrid assay for proteins that interacted with  
 CC TAK1. The 5' terminus was identified by 5'RACE. 2 Different  
 CC clones were sequenced, with cytosine and adenine (see AAT91178) as  
 CC the 185th nucleotide, respectively, and deposited as FERM BP-5599  
 CC and FERM BP-5508, respectively. Also claimed are: isolated DNA  
 CC encoding a protein modified by a substitution, deletion and/or  
 CC addition of 1 or more amino acids of the 504-residue TAB1 sequence;  
 CC (2) DNA which can hybridise with the 1560 bp nucleic acid sequence;  
 CC (3) isolated DNA encoding a protein comprising amino acids 21-579  
 CC or 437-504 of the 504 TAB1 sequence; (4) DNA encoding a fusion  
 CC protein comprising an above protein or polypeptide; (5) expression  
 CC vector comprising an above DNA; and (6) host cell, preferably a  
 CC mammalian or yeast cell, transformed by the expression vector.  
 CC Cells expressing TAB1 and TAK1 can be used to screen for TGF-beta  
 CC signalling pathway inhibitors by contacting the cells with a test  
 CC compound, and measuring the TAK1 kinase activity.  
 XX  
 SQ Sequence 1560 BP; 332 A; 469 C; 480 G; 279 T; 0 other;

Alignment Scores:  
 Pred. No.: 2,426-35 Length: 1560  
 Score: 359.00 Matches: 68  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 18 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x AAT91175 (1-1560)

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
 |||||  
 Db 1338 CAAGCCCGACCTTAACCTTCAGTCCACCAACAGCGACGACGAGCAGCTCCAGC 1397  
 |||||

Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40  
 |||||

```

Db 1398 TCTGACGAGGAGGCTCTCCGCTCCGCCGCCGACCTCGCTCCCGCTGCGAGACGCT 1457
Qy 41 ArgValAGluProTyrValAspPheAlaGluPheTyrArgLeuTyrSerValAspHisGly 60
Db 1458 CGGTGTGAGCCCTATGTGACTTCTGTGAGTTTACCGCTCTGAGCGGTGACCATGAC 1517
Qy 61 GluGlnSerValAlaThrAlaPro 68
Db 1518 GAGCAGAGCGTGTGTCAGACGACCG 1541

RESULT 3
AAT91178
ID AAT91178 standard; cDNA; 1560 BP.
XX
XX AAT91178;
AC
XX
XX 14-APR-1998 (first entry)
XX
XX Human TAB1 (TAK1 binding protein) cDNA.
DE
XX TAB1; TAK1 binding protein; transforming growth factor-beta;
KW signal transduction; human; ds.
XX
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT CDS 30..1544
FT /*tag= a
FT variation 185
FT /*tag= b
FT /note= "another clone has cytosine at position 185,
FT with codon AGA (Arg) altered to AGC (Ser)"
XX
XX EP803571-A2.
XX
XX 29-OCT-1997.
XX
XX 24-APR-1997; 97EP-0302808.
XX
XX 20-NOV-1996; 96US-0752891.
XX 24-APR-1996; 96JP-0126282.
XX 28-OCT-1996; 96JP-0300856.
XX
XX (UENOC/) UENOC N.
XX
XX Matsumoto K, Nishida E,
XX
XX WPI; 1997-515318/48.
XX P-PSDB; AAW26707.
XX
XX DNA encoding TAK1 binding protein TAB1 - member of transforming
XX growth factor beta receptor signal production pathway, which
XX activates TAK-1 kinase activity upon binding
XX
XX Example 5; Page 19-21; 30pp; English.
XX
XX This cDNA clone codes for human TAB1 (see AAW26707), a novel member
XX of the transforming growth factor-beta receptor signal transduction
XX pathway, which activates TAK-1 kinase activity upon binding. To
XX obtain the full-length TAB1 coding sequence, a human kidney cDNA
XX library was screened using as a probe a partial TAB1 cDNA obtained
XX from a yeast two-hybrid assay for proteins that interacted with
XX TAK1. The 5' terminus was identified by 5'RACE. 2 Different
XX clones were sequenced, with cytosine (see AAT91175) and adenine as
XX the 185th nucleotide, respectively, and deposited as FERM BP-5599
XX and FERM BP-5508, respectively. Also claimed are: isolated DNA
XX encoding a protein modified by a substitution, deletion and/or
XX addition of 1 or more amino acids of the 504-residue TAB1 sequence;
XX (2) DNA which can hybridise with the 1560 bp nucleic acid sequence;
XX (3) isolated DNA encoding a protein comprising amino acids 21-579
XX or 437-504 of the 504 TAB1 sequence; (4) DNA encoding a fusion
XX protein comprising an above protein or polypeptide; (5) expression
XX vector comprising an above DNA; and (6) host cell, preferably a

```

```

CC mammalian or yeast cell, transformed by the expression vector.
CC Cells expressing TAB1 and TAK1 can be used to screen for TGF-beta
CC signalling pathway inhibitors by contacting the cells with a test
CC compound, and measuring the TAK1 kinase activity.
XX
SQ Sequence 1560 BP; 333 A; 468 C; 480 G; 279 T; 0 other;

Alignment Scores:
Pred. No.: 2,42e-35 Length: 1560
Score: 359.00 Matches: 68
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0

US-09-830-144-4_copy_437_504 (1-66) x AAT91178 (1-1560)
Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisGlnSerSerSerSer 20
Db 1338 CAAAGCCGACCTTAACTGACAGTCCACACGACGACGACGACGACGCTTCAGC 1397
Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGlnAspGly 40
Db 1398 TCTGACGAGGAGGCTCTCCGCTCCGCCGCCGACCTCGCTCCCGCTGCGAGACGCT 1457
Qy 41 ArgValAGluProTyrValAspPheAlaGluPheTyrArgLeuTyrSerValAspHisGly 60
Db 1458 CGGTGTGAGCCCTATGTGACTTCTGTGAGTTTACCGCTCTGAGCGGTGACCATGAC 1517
Qy 61 GluGlnSerValAlaThrAlaPro 68
Db 1518 GAGCAGAGCGTGTGTCAGACGACCG 1541

RESULT 4
AAK56278
ID AAK56278 standard; DNA; 1560 BP.
XX
XX AAK56278;
AC
XX
XX 21-JUL-1999 (first entry)
XX
XX Human TAB1 encoding DNA.
DE
XX
XX Human; TAB1; screening; inhibition; TGF-beta;
XX transforming growth factor beta; ss.
XX
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT CDS 30..1544
FT /*tag= a
FT
XX
XX WO9921010-A1.
XX
XX 29-APR-1999.
XX
XX 22-OCT-1998; 98WO-JP04796.
XX
XX 22-OCT-1997; 97JP-0290188.
XX
XX (CHUS ) CHUGAI SEIYAKU KK.
XX
XX Ohtomo T, Ono K, Tsuchiya M;
XX
XX WPI; 1999-312645/26.
XX P-PSDB; AAY09541.
XX
XX Screening for TGF-beta inhibitory substances, which are useful as
XX drugs for treatment of diseases relating to its disorder
XX
XX Example 1; Page 143-147; 195pp; Japanese.
XX
XX A method has been developed for screening for substances which inhibit

```

CC the binding of TAK1 polypeptide to TAB1 polypeptide. The method  
 CC comprises: (a) contacting the polypeptide in the presence of a sample;  
 CC and (b) detecting the amount of bound polypeptide, in which the sample  
 CC can be pre-mixed with TAK1 or TAB1 polypeptide first. The transforming  
 CC growth factor (TGF)-beta inhibitory substances can be used in drugs for  
 CC indications e.g. as TGF-beta signal transduction inhibitors or  
 CC activators, or extracellular matrix protein production enhancement  
 CC inhibitors, or activators, or cell proliferation prevention inhibitors or  
 CC activators, or monocyte migration inhibitors or activators, or  
 CC physiological activity induction inhibitors or activators, or  
 CC immunosuppression inhibitors or activators, or amyloid beta protein  
 CC precipitation inhibitors or activators, and such substances can also be  
 CC inhibitors of the TAK1 polypeptide function, particularly kinase  
 CC activity. The present sequence encodes human TAB1.

XX SQ Sequence 1560 BP; 332 A; 469 C; 480 G; 279 T; 0 other;

Alignment Scores: 2,42e-35 Length: 1560  
 Pred. No.: 359.00 Matches: 68  
 Score: 359.00  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 20 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x AAX56278 (1-1560)

Oy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
 Db 1338 CAAGCGCGACCTTAACCTCGAGTCCACCCACACGACGACGAGGAGCGTCCAGC 1397  
 Oy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40  
 Db 1398 TCTGACGAGGAGGCTTTCCGCTCCCGCGCCGCCACCTCGCTCCCGCCCTGGCGAGGCGGT 1457  
 Oy 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
 Db 1458 CGTGTGAGCCCTATGTGGACTTTGCTGAGTTTACCGGCTCTGGAGGCGTGGACCATGGC 1517  
 Oy 61 GluGlnSerValThrAlaPro 68  
 Db 1518 GAGCAGAGCGTGGTGACAGCACCG 1541

RESULT 5

AAA39106  
 ID AAA39106 standard; DNA; 1560 BP.

XX AC AAA39106;

XX DT 04-SEP-2000 (first entry)

XX DE Human TAB-1 nucleotide sequence SEQ ID NO:3.

XX KW Human; TAK-1; TAB-1; mitogen activated protein kinase; MAPK;  
 KW screening; signal transduction; inhibition; inflammatory cytokine;  
 KW IL-1; interleukin 1; TNF; tumour necrosis factor; inflammation;  
 KW antiinflammatory; suppression; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 XX CDS 30..1544  
 XX FT /\*tag= a  
 XX FT /product= "TAB-1"

XX PN WO2000023610-A1.

XX PD 27-APR-2000.

XX PF 21-OCT-1999; 99WO-JP05817.

XX PR 21-OCT-1998; 98JP-0299962.

PA (CHUS ) CHUGAI SEIYAKU KK.  
 XX PI Tsuchiya M, Ohtomo T, Sugamata Y, Matsumoto K;  
 XX DR WPI; 2000-339707/29.  
 XX DR P-FSDB; AAV91001.

XX Method for screening inhibitors of TAK1 signal transduction for  
 PT suppression of inflammatory cytokine production and use as  
 PT antiinflammatory agents

XX PS Disclosure; Page 85-90; 100pp; Japanese.

XX The present invention describes a method for screening compounds for  
 CC inhibition of inflammatory cytokine signal transduction by contacting  
 CC the sample with TAK1 and its receptor TAB1 and selecting for inhibition  
 CC of TAK1/TAB1 binding. Also described is a method for screening compounds  
 CC for inhibition of inflammatory cytokine signal transduction in which the  
 CC inhibition of TAK1 phosphorylation is selected for; and drug  
 CC compositions for the treatment of inflammatory disorders containing as  
 CC active component an inflammatory cytokine signal transduction inhibitor.  
 CC TAK1 is an essential component of the signalling process which results  
 CC in release of inflammatory cytokines such as interleukin-1 (IL-1),  
 CC IL-10, tumour necrosis factor (TNF) and IL-6. The methods can be used  
 CC for the selection of effective antiinflammatory agents. The present  
 CC sequence encodes human TAB-1, which is used in the exemplification of the  
 CC present invention.

XX SQ Sequence 1560 BP; 332 A; 469 C; 480 G; 279 T; 0 other;

Alignment Scores:

Pred. No.: 2,42e-35 Length: 1560  
 Score: 359.00 Matches: 68  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x AAA39106 (1-1560)

Oy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
 Db 1338 CAAGCGCGACCTTAACCTCGAGTCCACCCACACGACGACGAGGAGCGTCCAGC 1397

Oy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40  
 Db 1398 TCTGACGAGGAGGCTTTCCGCTCCCGCGCCGCCACCTCGCTCCCGCCCTGGCGAGGCGGT 1457

Oy 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
 Db 1458 CGTGTGAGCCCTATGTGGACTTTGCTGAGTTTACCGGCTCTGGAGGCGTGGACCATGGC 1517

Oy 61 GluGlnSerValThrAlaPro 68

Db 1518 GAGCAGAGCGTGGTGACAGCACCG 1541

RESULT 6

AAA56310  
 ID AAA56310 standard; DNA; 1568 BP.

XX AC AAX56310;

XX DT 21-JUL-1999 (first entry)

XX DE Human TAB1 encoding DNA SEQ ID NO:42.

XX KW Human; TAB1; TAK1; screening; inhibition; TGF-beta;  
 KW transforming growth factor beta; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 11..1552

```
PT XX /+tag= a
XX PN MO9921010-A1.
XX PD 29-APR-1999.
XX PF 22-OCT-1998; 98WO-JP04796.
XX PR 22-OCT-1997; 97JP-0290188.
XX PA (CHUS ) CHUGAI SEIYAKU KK.
XX PI Ohtomo T, Ono K, Tsuchiya M;
XX DR WPI; 1999-312645/26.
XX DR P-PSDB; AAY09550.
PT PT Screening for TGF- beta inhibitory substances, which are useful as
XX PT drugs for treatment of diseases relating to its disorder
XX PS Example 13; Page 182-186; 195pp; Japanese.
XX CC A method has been developed for screening for substances which inhibit
XX CC the binding of TAK1 polypeptide to TAB1 polypeptide. The method
XX CC comprises: (a) contacting the polypeptide in the presence of a sample;
XX CC and (b) detecting the amount of bound polypeptide, in which the sample
XX CC can be pre-mixed with TAK1 or TAB1 polypeptide first. The transforming
XX CC growth factor (TGF)-beta inhibitory substances can be used in drugs for
XX CC indications e.g. as TGF-beta signal transduction inhibitors or
XX CC activators, or extracellular matrix protein production enhancement
XX CC inhibitors, or activators, or cell proliferation prevention inhibitors or
XX CC activators, or monocyte migration inhibitors or activators, or
XX CC physiological activity induction inhibitors or activators, or
XX CC immunosuppression inhibitors or activators, and such substances can also be
XX CC precipitation inhibitors or activators, and such substances can also be
XX CC inhibitors of the TAK1 polypeptide function, particularly kinase
XX CC activity. The present sequence encodes human TAB1.
XX SQ Sequence 1568 BP; 339 A; 472 C; 477 G; 280 T; 0 other;

Alignment Scores:
Pred. No.: 2,44e-35 Length: 1568
Score: 359.00 Matches: 68
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-830-144-4_COPY_437_504 (1-68) x AAX56310 (1-1568)
QY 1 GlnSerProThrLeuThriLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20
Db 1346 CAAAGCCCGACCTTAACCTTCGACGTCACCAACGACACGACAGACAGCTCCAGC 1405
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGly 40
Db 1406 TCTGACGAGAGCGCTCTCCGCTCCGCGCCGCACTCCGCTCCGCTGCGGAGAGAGGCT 1465
QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuThrSerValAspHisGly 60
Db 1466 CGTGTGAGCCCTTAATGTGAGACTTTCCTGAGCTTTACCGGCTCTGAGAGCTGACCAAGGC 1525
QY 61 GlnSerProThrLeuThriLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 68
Db 1526 GAGCAGAGCGGTGTGACGACGACCG 1549
RESULT 7
ID AAX56282
XX AAX56282 standard; DNA; 1569 BP.
AC AAX56282;
XX 21-JUN-1999 (first entry)
DT
```

```
XX XX Human TAB1-FLAG encoding DNA.
DE DE
XX XX Human; TAB1; screening; inhibition; TGF-beta;
XX KM transforming growth factor beta; ss.
XX OS Homo sapiens.
XX OS Synthetic.
FH FH Key Location/Qualifiers
FT FT CDS 7..1560
XX /+tag= a
XX PN MO9921010-A1.
XX PD 29-APR-1999.
XX PF 22-OCT-1998; 98WO-JP04796.
XX PR 22-OCT-1997; 97JP-0290188.
XX PA (CHUS ) CHUGAI SEIYAKU KK.
XX PI Ohtomo T, Ono K, Tsuchiya M;
XX DR WPI; 1999-312645/26.
XX DR P-PSDB; AAY09546.
PT PT Screening for TGF- beta inhibitory substances, which are useful as
XX PT drugs for treatment of diseases relating to its disorder
XX PS Example 1; Page 159-163; 195pp; Japanese.
XX CC A method has been developed for screening for substances which inhibit
XX CC the binding of TAK1 polypeptide to TAB1 polypeptide. The method
XX CC comprises: (a) contacting the polypeptide in the presence of a sample;
XX CC and (b) detecting the amount of bound polypeptide, in which the sample
XX CC can be pre-mixed with TAK1 or TAB1 polypeptide first. The transforming
XX CC growth factor (TGF)-beta inhibitory substances can be used in drugs for
XX CC indications e.g. as TGF-beta signal transduction inhibitors or
XX CC activators, or extracellular matrix protein production enhancement
XX CC inhibitors, or activators, or cell proliferation prevention inhibitors or
XX CC activators, or monocyte migration inhibitors or activators, or
XX CC physiological activity induction inhibitors or activators, or
XX CC immunosuppression inhibitors or activators, or amyloid beta protein
XX CC precipitation inhibitors or activators, and such substances can also be
XX CC inhibitors of the TAK1 polypeptide function, particularly kinase
XX CC activity. The present sequence encodes TAB1-FLAG from an example of
XX CC the present invention.
XX SQ Sequence 1569 BP; 343 A; 466 C; 476 G; 284 T; 0 other;

Alignment Scores:
Pred. No.: 2,44e-35 Length: 1569
Score: 359.00 Matches: 68
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-830-144-4_COPY_437_504 (1-68) x AAX56282 (1-1569)
QY 1 GlnSerProThrLeuThriLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20
Db 1315 CAAAGCCCGACCTTAACCTTCGACGTCACCAACGACACGACAGACAGCTCCAGC 1374
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGly 40
Db 1375 TCTGACGAGAGCGCTCTCCGCTCCGCGCCGCACTCCGCTCCGCTGCGGAGAGAGGCT 1434
QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuThrSerValAspHisGly 60
Db 1435 CGTGTGAGCCCTTAATGTGAGACTTTCCTGAGCTTTACCGGCTCTGAGAGCTGACCAAGGC 1494
```

QY 61 GluGlnSerValThrAlapro 68  
DB 1495 GAGCAGCGTGGTGACGACCG 1518  
RESULT 8  
ID ABA20494 standard; DNA; 16877 BP.  
XX ABA20494;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Human nervous system related polynucleotide SEQ ID NO 12825.  
XX Human; neotropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antiskling; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01334.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225477.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246529.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.

PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JUN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -  
XX  
XX Disclosure; SEQ ID NO 12825; 1701bp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins  
CC (ABAI4678-ABH18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WHO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 16877 BP; 3563 A; 4508 C; 4516 G; 4290 T; 0 other;

Alignment Scores:  
Pred. No.: 5 556-34 Length: 16877  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 22 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x ABA20494 (1-16877)

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAanThrHisThrGlnSerSerSerSer 20  
Db 15011 CAAAGCCGAGCCTTAACCTGAGCTCACCAACAGCAGCAGAGAGAGAGCTCCAGC 15070

Qy 21 SerAspGlyGlyLeuPheArgSerARGPProAlaHisSerLeuProGlyGlyAspGly 40  
Db 15071 TCTACGAGAGCCTCTTCCGCTCCGCGCCGCCCACTCGCTCCGCTCGGAGAGAGCGT 15130

Qy 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
Db 15131 CGGTGTGAGCCCTTATGTAGTGCATTTGCTGAGTTTACCGCTCGAGCGTGCATGAC 15190

Qy 61 GluGlnSerValValThrAlaPro 68  
Db 15191 GAGCAGAGCGCTGTGTACACAGACCG 15214

RESULT 9  
AAL36984  
ID AAL36984 standard; DNA; 16877 BP.  
XX  
AC AAL36984;  
XX  
DT 08-JUN-2002 (first entry)  
XX  
XX Human musculoskeletal system related polynucleotide SEQ ID NO 3349.  
DE  
XX  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anticancer;  
KW vulnegracy; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200155367-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01338.  
XX  
PR 31-JUN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184668.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190706.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0224521.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0226279.  
PR 18-AUG-2000; 2000US-0226281.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227109.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.





QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGly 40  
 DB 15071 TCTGACCGAGGCGCTCTCCGCTCCGCGCGGCACTCCGCTCCGCGCGAGGAGACCGT 15130  
 QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
 DB 15131 CGTGTGAGGCCCTATGAGGACTTGGAGTTTACCGGCTCTGAGAGCGTGAGCCATGGC 15190  
 QY 61 GluGlnSerValValThrAlaPro 68  
 DB 15191 GAGCAGAGCGGTGTGACAGCACCG 15214  
 RESULT 10  
 AAF15895  
 ID AAF15895 standard; cDNA; 696 BP.  
 AC AAF15895;  
 XX  
 DT 13-MAR-2001 (first entry)  
 DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:330.  
 XX  
 KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
 KW neuroprotective; cytoskeletal; immunomodulatory; muscular;  
 KW vulnereary; gastrointestinal; nephrotoxic; antiinfective; gynaecological;  
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
 KW wound; infectious disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO20005174-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05988.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2000-587513/55.  
 DR P-PSDB; AAB56692.  
 XX  
 PT Prostate cancer associated gene sequences, referred to as prostate  
 PT cancer antigens, useful for treatment, prevention, and diagnosis of  
 PT disorders such as prostate cancer -  
 XX  
 PS Claim 1; Page 837; 2338pp; English.  
 XX  
 CC AAF1566 to AAF16505 encode the human prostate cancer associated  
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
 CC The prostate cancer antigens can have neuroprotective, cytoskeletal,  
 CC cardiotropic, immunomodulatory, muscular, vulnereary, gastrointestinal,  
 CC nephrotoxic, antiinfective, gynaecological and antibacterial activities,  
 CC and can be used in gene therapy. The prostate cancer antigen  
 CC polynucleotides may be used for detection of prostate cancer, chromosome  
 CC identification, as chromosome markers, and for numerous other diagnostic  
 CC or research purposes. The prostate cancer antigens may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
 CC AAB57303 represent sequences used in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 696 BP; 143 A; 237 C; 187 G; 120 T; 9 other;  
 Alignment Scores: 6 36e-35 Length: 696  
 Pred. No.: 352.00 Matches: 67  
 Score:

Percent Similarity: 98.53% Conservativeness: 0  
 Best Local Similarity: 98.53% Mismatches: 1  
 Query Match: 98.05% Indels: 0  
 DB: 21 Gaps: 0  
 US-09-830-144-4\_COPY\_437\_504 (1-68) x AAF15895 (1-696)  
 QY 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
 DB 49 CAAAGCCGACCTTAACCTTGACAGTCCACCAACAGCAGCAGGAGGAGGAGGAGGAGG 108  
 QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGly 40  
 DB 109 TCTRACGAGGCGCTCTCCGCTCCGCGCGGCACTCCGCTCCGCGCGAGGAGCGGT 168  
 QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
 DB 169 CGTGTGAGGCCCTATGAGGACTTGGAGTTTACCGGCTCTGAGAGCGTGAGCCATGGC 228  
 QY 61 GluGlnSerValValThrAlaPro 68  
 DB 229 GAGCAGAGCGGTGTGACAGCACCG 252  
 RESULT 11  
 ID AABN61139/c  
 AC AABN61139;  
 XX  
 DT 28-JUN-2002 (first entry)  
 DE Human cancer related polynucleotide SEQ ID NO 1106.  
 XX  
 KW Human; cytoskeletal; gene expression; gene mapping; tissue profiling;  
 KW gene therapy; cancer; tumour; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200214500-A2.  
 XX  
 PD 21-FEB-2002.  
 XX  
 PF 16-AUG-2001; 2001WO-US25840.  
 XX  
 PR 16-AUG-2000; 2000US-226326P.  
 XX  
 PA (CHIR-) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;  
 PI Lamson G, Scott EM, Zhang G, Kaasam A, Bot D, Labat I;  
 XX  
 DR WPI; 2002-241905/29.  
 XX  
 CC New nucleic acid for producing a polypeptide, detecting differentially  
 CC expressed genes correlated with a cancerous state of a mammalian cell,  
 CC and inhibiting tumor growth -  
 XX  
 PS Claim 1; SEQ ID NO 1106; 883pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)  
 CC with cytoskeletal activity. The polynucleotide is used to produce a  
 CC polypeptide, to detect differentially expressed genes correlated with a  
 CC cancerous state of a mammalian cell and to inhibit tumour growth. The  
 CC polynucleotide is used as a probe in mapping and tissue profiling. The  
 CC encoded polypeptide and antibodies to the polypeptide can also be used  
 CC for therapeutic and diagnostic purposes. The polynucleotide is useful for  
 CC gene therapy.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 519 BP; 152 A; 118 C; 114 G; 135 T; 0 other;



CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AA64197-AA64564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 990 BP; 215 A; 318 C; 287 G; 170 T; 0 other;

Alignment Scores:  
Pred. No.: 8.18 Length: 990  
Score: 74.50 Matches: 22  
Percent Similarity: 38.96% Conservative: 8  
Best Local Similarity: 28.57% Mismatches: 28  
Query Match: 20.75% Indels: 19  
DB: 23 Gaps: 2

US-09-830-144-4\_COPY\_437\_504 (1-68) x AA80813 (1-990)

QY 2 SerProThrIleuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSer 21  
DB 412 GCGCCACACGCCCGGATGCGGACCGCGCCACCCACAGTCCTGGGAGACTTCA 471  
QY 22 AspGlyGlyLeuPheArgProAlaHisSerLeuProGlyGlnAspGly--- 40  
DB 472 GATGGG-----AGGAACAGGTGCCCCCAGATGCTCGAAAGTAACCATGGCCGC 522  
QY 41 -----ArgValGlnProTyr 45  
DB 523 CCGATGGCAGCTCAGCTGTTTGGGATATTATTGCGATCGCAGAGTGAAGAGAAC 582  
QY 46 ValAspPheAlaGlnPheTyrArgLeuTipSerValAspHisGlyGln 62  
DB 583 ATTCAGATGTGAAGACTTCAGAACGTGGACCTCGTTCGTGAGAGAG 633

RESULT 14  
AAQ49396/C  
ID AAQ49396 standard; DNA; 1108 BP.  
XX  
AC AAQ49396;  
XX  
DT 27-APR-1994 (first entry)  
XX  
DE Class II AP endonuclease.  
XX  
KW Apurinic/apyrimidinic; recombinant; heat stable; ss.  
XX  
OS Thermus thermophilus.  
XX  
FH Key Location/Qualifiers  
FT CDS 114..926  
FT /\*tag= a  
XX  
PN MO9320191-A.  
XX  
PD 14-OCT-1993.  
XX  
PF 30-MAR-1993; 93WO-US03035.  
XX  
PR 31-MAR-1992; 92US-0860861.  
PR 16-APR-1992; 92US-0869306.  
XX

PA (ABBO ) ABBOTT LAB.  
PA (CHAD ) CHAKAKI DYEING KK.  
XX  
PI Backman KC, McLean PA;  
XX  
XX WPI; 1993-336899/42.  
DR P-PSDB; AAR42817.  
XX  
XX New recombinant DNA encoding heat stable AP endonuclease - and  
PT derived vectors, host cells and proteins, useful for nucleic acid  
PT amplification by ligase chain reaction  
XX  
PS Claim 4; Fig 1A; 42pp; English.  
XX  
CC The sequence is that encoding Thermus thermophilus class II AP  
CC (apurinic/apyrimidinic) endonuclease which retains activity when  
CC subjected to elevated temperatures for the time necessary to  
CC effect denaturation of double-stranded nucleic acids.  
XX  
SQ Sequence 1108 BP; 170 A; 379 C; 393 G; 166 T; 0 other;

Alignment Scores:  
Pred. No.: 22.6 Length: 1108  
Score: 71.50 Matches: 15  
Percent Similarity: 70.37% Conservative: 4  
Best Local Similarity: 55.56% Mismatches: 7  
Query Match: 19.92% Indels: 1  
DB: 14 Gaps: 1

US-09-830-144-4\_COPY\_437\_504 (1-68) x AAQ49396 (1-1108)

QY 18 SerSerSerSerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGly 37  
DB 203 AGCGGTAGCGCCGAGGCGGCGGTCTCTCCACCGCCGCCACGCGCTTTTCCCGC 144  
QY 38 GlnAspGlyArgValGlnPro 44  
DB 143 --GATGGAAAGGTGAACCC 126

RESULT 15  
ABK78928  
ID ABK78928 standard; DNA; 996 BP.  
XX  
AC ABK78928;  
XX  
DT 13-AUG-2002 (first entry)  
XX  
DE Bacillus clausii genomic sequence tag (GST) #1771.  
XX  
XX Differential gene expression; genomic sequenced tag; GST;  
KW altered culture condition; environmental stress;  
XX physiological provocation; ds.  
XX  
OS Bacillus clausii.  
XX  
PN WO200229113-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 05-OCT-2001; 2001WO-US31437.  
XX  
PR 06-OCT-2000; 2000US-0680598.  
PR 27-MAR-2001; 2001US-279526P.  
XX  
PA (NOVO ) NOVOZYMES BIOTECH INC.  
PA (NOVO ) NOVOZYMES AS.  
XX  
PI Berka R, Clausen IG;  
XX  
DR WPI; 2002-416684/44.  
XX  
PT Monitoring differential expression of several genes in first Bacillus  
PT cell relative to expression of same genes in one or more second



```

; EARLIER APPLICATION NUMBER: US No. 608712209/070,060
; EARLIER FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5372
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-357-746-1

Alignment Scores:
Pred. No.: 119          Length: 5372
Score: 67.50           Matches: 23
Percent Similarity: 47.30% Conservative: 12
Best Local Similarity: 31.08% Mismatches: 20
Query Match: 18.80%      Indels: 19
DB: 3                  Gaps: 3

US-09-830-144-4_COPY_437_504 (1-68) x US-09-357-746-1 (1-5372)
QY 2 SerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSer 21
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 843 TCTAGTACAGGCTCTCTGCCCCACAAATACAAATACATCTGAGAGCAACA 902
QY 22 AspGly-----GlyLeuPheArgSerArgProIahis--- 32
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 903 TCTGATTAAATATCTCTTACTATATCTGAGGCTCAGGCCCTGAGCATTAATCTCT 962
QY 33 -----SerLeuProProGlyLuaspGlyArgValGluProTyrValAspPhe 48
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 963 GTAACACAGCTCCCTTCCACCTGCTGGAGAGCAGAGATGAC----- 1007
QY 49 AlaGluPheTyrArgLeuTyrPseValAspHisGlyGluGln 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1008 --CAGCAGGCGAGGTTACTATGATGATCATGTTGAGAAA 1046

RESULT 15
US-09-152-060-23
; Sequence 23, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P1 US
; CURRENT APPLICATION NUMBER: US/09/152,060
; EARLIER FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/US98/04858
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: 60/040,762
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/040,710
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/050,934
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,100
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,357
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,189
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/057,765
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/068,368
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-152-060-23
```

```

Alignment Scores:
Pred. No.: 48.5          Length: 1101
Score: 63.50           Matches: 27
Percent Similarity: 41.43% Conservative: 2
Best Local Similarity: 38.57% Mismatches: 30
Query Match: 17.69%      Indels: 11
DB: 4                  Gaps: 3

US-09-830-144-4_COPY_437_504 (1-68) x US-09-152-060-23 (1-1101)
QY 7 LeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSerAsp----- 22
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 576 CTGACAGCGGCGGACCTGCGGGGTACAGCAGACAGCTGCGAGCCTCCTGGACCC 635
QY 23 -----GlyGlyLeuPheArgSerArgProIahisSerLeuProGlyGluaspGly 40
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 636 CGAGCGGACCGCTCTTCAGGCTCGGCCCGGGGACTCTCTGCCATCCCGAGAGACCCGC 695
QY 41 ArgValGluProTyr-----ValAspPheAlaGluPheTyrArgLeuTyrPse 56
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 696 TACGTGAGCTGTATGTGTCGTGACATGACAGAGTTCAGATCTGAGGAGCGAAGCA 755
QY 57 ---ValAspHisGlyGluGlnSerValVal 65
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 756 GCCGTGCTATCGGGTGGTGTGAGGTGTG 785

Search completed: December 10, 2002, 07:27:27
Job time : 72 secs
```

```
Db 694 TCTAGTACAGGCTCTCTCCGCCGACAAATACAAATACATCTCGAAGGACACA 753
Qy 22 AspGly-----GlyLeuPheArgSerArgProAlaHis--- 32
|||
Db 754 TCTGGATTAAATTCCTCTTACTATATCTGGAGGCTCAGGCCCTAGGCCATTAAATCCT 813
|||
Qy 33 -----SerLeuProGlyGluAspGlyArgValGluProTyrValAspPhe 48
|||
Db 814 GTAACCTCAAGCTCCCTTGCACCTGGTGGGAGCAGAGTGGAC----- 858
|||
Qy 49 AlaGluPheTyrArgLeuTrpSerValAspHisGlyGluGln 62
|||
Db 859 ---CAGCAGCGGCGAGTTTACTATGTAGATCATCTTGAGAAA 897
|||
RESULT 12
US-09-357-746-2
; Sequence 2, Application US/09357746
; Patent No. 6087122
; GENERAL INFORMATION:
; APPLICANT: ZENECA Limited
; TITLE OF INVENTION: HUMAN E3 UBIQUITIN PROTEIN LIGASE
; FILE REFERENCE: PHM.70312.NI
; CURRENT APPLICATION NUMBER: US/09/357,746
; EARLIER FILING DATE: 1999-07-21
; EARLIER FILING DATE: 1998-02-05
; EARLIER FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2559
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-357-746-2

Alignment Scores:
Pred. No.: 45.6 Length: 2559
Score: 67.50 Matches: 23
Percent Similarity: 47.30% Conservative: 12
Best Local Similarity: 31.08% Mismatches: 20
Query Match: 18.80% Indels: 19
DB: 3 Gaps: 3

US-09-830-144-4_COPY_437_504 (1-68) x US-09-357-746-2 (1-2559)
Qy 2 SerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 21
|||
Db 694 TCTAGTACAGGCTCTCTCCGCCGACAAATACAAATACATCTCGAAGGACACA 753
|||
Qy 22 AspGly-----GlyLeuPheArgSerArgProAlaHis--- 32
|||
Db 754 TCTGGATTAAATTCCTCTTACTATATCTGGAGGCTCAGGCCCTAGGCCATTAAATCCT 813
|||
Qy 33 -----SerLeuProGlyGluAspGlyArgValGluProTyrValAspPhe 48
|||
Db 814 GTAACCTCAAGCTCCCTTGCACCTGGTGGGAGCAGAGTGGAC----- 858
|||
Qy 49 AlaGluPheTyrArgLeuTrpSerValAspHisGlyGluGln 62
|||
Db 859 ---CAGCAGCGGCGAGTTTACTATGTAGATCATCTTGAGAAA 897
|||
RESULT 13
US-09-070-060-1
; Sequence 1, Application US/09070060
; Patent No. 5976849
; GENERAL INFORMATION:
; APPLICANT: Husted, Carolyn M.
; APPLICANT: Childyal, Namit
; TITLE OF INVENTION: Human E3 Ubiquitin Protein
; FILE REFERENCE: PHM.70312.NI
; CURRENT APPLICATION NUMBER: US/09/357,746
; EARLIER FILING DATE: 1999-07-21
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: ZENECA Pharmaceuticals, Inc.
; STREET: 1800 Concord Pike
; CITY: Wilmington
; STATE: DE
; COUNTRY: USA
; ZIP: 19850-5437
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,060
; FILING DATE: 30-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/073,839
; FILING DATE: 05-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Higgins, Patrick H
; REGISTRATION NUMBER: 39,709
; REFERENCE/DOCKET NUMBER: PHM.70312
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302.886.4889
; TELEFAX: 302.886.8221
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-070-060-1

Alignment Scores:
Pred. No.: 119 Length: 5359
Score: 67.50 Matches: 23
Percent Similarity: 47.30% Conservative: 12
Best Local Similarity: 31.08% Mismatches: 20
Query Match: 18.80% Indels: 19
DB: 3 Gaps: 3

US-09-830-144-4_COPY_437_504 (1-68) x US-09-070-060-1 (1-5359)
Qy 2 SerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 21
|||
Db 848 TCTAGTACAGGCTCTCTCCGCCGACAAATACAAATACATCTGAAGGACACA 907
|||
Qy 22 AspGly-----GlyLeuPheArgSerArgProAlaHis--- 32
|||
Db 908 TCTGGATTAAATTCCTCTTACTATATCTGGAGGCTCAGGCCCTAGGCCATTAAATCCT 967
|||
Qy 33 -----SerLeuProGlyGluAspGlyArgValGluProTyrValAspPhe 48
|||
Db 968 GTAACCTCAAGCTCCCTTGCACCTGGTGGGAGCAGAGTGGAC----- 1012
|||
Qy 49 AlaGluPheTyrArgLeuTrpSerValAspHisGlyGluGln 62
|||
Db 1013 ---CAGCAGCGGCGAGTTTACTATGTAGATCATCTTGAGAAA 1051
|||
RESULT 14
US-09-357-746-1
; Sequence 1, Application US/09357746
; Patent No. 6087122
; GENERAL INFORMATION:
; APPLICANT: ZENECA Limited
; TITLE OF INVENTION: HUMAN E3 UBIQUITIN PROTEIN LIGASE
; FILE REFERENCE: PHM.70312.NI
; CURRENT APPLICATION NUMBER: US/09/357,746
; EARLIER FILING DATE: 1999-07-21
; EARLIER FILING DATE: 1998-02-05
```



```
QY 41 AtgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60
Db 1458 CGTGTGAGCCCTATGTGACCTTTGCTGAGTTTACCGCTCTGGAGCGTGGACCATGCG 1517

QY 61 GluGlnSerValValThrAlaPro 68
Db 1518 GAGCAGAGCGTGTGACAGCACCG 1541

RESULT 7
US-09-529-279-1
; Sequence 1, Application US/09529279
; Patent No. 6451617
; GENERAL INFORMATION:
; APPLICANT: ONO, KOICHIRO
; APPLICANT: OHTOMO, TOSHIHIKO
; APPLICANT: TSUCHIYA, MASAYUKI
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES
; FILE REFERENCE: 053466/0278
; CURRENT APPLICATION NUMBER: US/09/529,279
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: PCT/JP98/04796
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: JP 9/290188
; PRIOR FILING DATE: 1997-10-22
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (30)..(1541)
US-09-529-279-1
Alignment Scores:
Pred. No.: 6,56e-36 Length: 1560
Score: 359.00 Matches: 68
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
US-09-830-144-4_COPY_437_504 (1-68) x US-09-529-279-1 (1-1560)

QY 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20
Db 1338 CAAAGCCCGACCTTAACCTGCGAGTCCACCAACAGCACGACGAGCAGCAGCTCCAGC 1397

QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40
Db 1398 TCTGACGAGGCGCTCTCCGCTCCCGCCGCCCACTCGCTCCGCTGGCGAGGACGCT 1457

QY 41 AtgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60
Db 1458 CGTGTGAGCCCTATGTGACCTTTGCTGAGTTTACCGCTCTGGAGCGTGGACCATGCG 1517

QY 61 GluGlnSerValValThrAlaPro 68
Db 1518 GAGCAGAGCGTGTGACAGCACCG 1541

RESULT 8
US-09-529-279-42
; Sequence 42, Application US/09529279
; Patent No. 6451617
; GENERAL INFORMATION:
; APPLICANT: ONO, KOICHIRO
; APPLICANT: OHTOMO, TOSHIHIKO
; APPLICANT: TSUCHIYA, MASAYUKI
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES
; FILE REFERENCE: 053466/0278
; CURRENT APPLICATION NUMBER: US/09/529,279
; PRIOR FILING DATE: 2000-04-11
```

```
; PRIOR APPLICATION NUMBER: PCT/JP98/04796
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: JP 9/290188
; PRIOR FILING DATE: 1997-10-22
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 1568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11)..(1549)
US-09-529-279-42
Alignment Scores:
Pred. No.: 6,6e-36 Length: 1568
Score: 359.00 Matches: 68
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
US-09-830-144-4_COPY_437_504 (1-68) x US-09-529-279-42 (1-1568)

QY 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20
Db 1346 CAAAGCCCGACCTTAACCTGCGAGTCCACCAACAGCACGACGAGCAGCTCCAGC 1405

QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40
Db 1406 TCTGACGAGGCGCTCTCCGCTCCCGCCGCCCACTCGCTCCGCTGGCGAGGACGCT 1465

QY 41 AtgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60
Db 1466 CGTGTGAGCCCTATGTGACCTTTGCTGAGTTTACCGCTCTGGAGCGTGGACCATGCG 1525

QY 61 GluGlnSerValValThrAlaPro 68
Db 1526 GAGCAGAGCGTGTGACAGCACCG 1549

RESULT 9
US-09-529-279-10
; Sequence 10, Application US/09529279
; Patent No. 6451617
; GENERAL INFORMATION:
; APPLICANT: ONO, KOICHIRO
; APPLICANT: OHTOMO, TOSHIHIKO
; APPLICANT: TSUCHIYA, MASAYUKI
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES
; FILE REFERENCE: 053466/0278
; CURRENT APPLICATION NUMBER: US/09/529,279
; CURRENT FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: PCT/JP98/04796
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: JP 9/290188
; PRIOR FILING DATE: 1997-10-22
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 1569
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1557)
US-09-529-279-10
Alignment Scores:
Pred. No.: 6,61e-36 Length: 1569
Score: 359.00 Matches: 68
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
```



NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/406,854  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,891  
FILING DATE: 20-NOV-1996  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1560 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1541  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 30..1541  
US-09-406-854-1  
Alignment Scores:  
Pred. No.: 6.56e-36 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
DB: 3  
US-09-830-144-4\_COPY\_437\_504 (1-68) x US-09-406-854-1 (1-1560)  
QY 1 GlnSerProThrLeuThrIleuGlnSerThrAsnThrHisThrGlnSerSerSerSerSer 20  
Db 1338 CAAAGCCCACTTAACCTTCGACGTCACCAACAGCAGCAGCAGCAGCAGCAGCAGCAGC 1397  
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyYgiuaSpGly 40  
Db 1398 TCTGACGAGAGGCTCTTCGCTCCGCGCCGCGCCACTGCTCCGCTGCGCGAGAGGAT 1457  
QY 41 ArgValGluProTyValAspPheAlaGluPheTyArgLeuTyPserValAspHisGly 60  
Db 1458 CGTGTGACCCCTATGTGGAGCTTCTGCTGAGTTTACCCCTCTGAGACGTCGACCATGGC 1517  
QY 61 GluGlnSerValValThrAlaPro 68  
Db 1518 GAGCAGAGCGGTGGTGACGACCGC 1541

RESULT 6  
US-09-406-854-5  
Sequence 5, Application US/09406854  
Patent No. 6140042  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihiko  
APPLICANT: NISHIDA, Betsuke  
TITLE OF INVENTION: TABI PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/406,854  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,891  
FILING DATE: 20-NOV-1996  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1560 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1541  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 30..1541  
US-09-406-854-5  
Alignment Scores:  
Pred. No.: 6.56e-36 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
DB: 3  
US-09-830-144-4\_COPY\_437\_504 (1-68) x US-09-406-854-5 (1-1560)  
QY 1 GlnSerProThrLeuThrIleuGlnSerThrAsnThrHisThrGlnSerSerSerSerSer 20  
Db 1338 CAAAGCCCACTTAACCTTCGACGTCACCAACAGCAGCAGCAGCAGCAGCAGCAGCAGC 1397  
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyYgiuaSpGly 40  
Db 1398 TCTGACGAGAGGCTCTTCGCTCCGCGCCGCGCCACTGCTCCGCTGCGCGAGAGGAT 1457

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/752,891  
FILING DATE: 20-NOV-1996  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1560 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1541  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 30..1541  
US-09-144-178-1

Alignment Scores:  
Pred. No.: 6.56e-36 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x US-09-144-178-1 (1-1560)

QY 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
DB 1338 CAAGACCCGACCTTAAACCTTCAGTCCACACACGACGACGAGCAGCTCCACG 1397  
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40  
DB 1398 TCTGACGAGGCGCTTCTCCGCTCCCGGCCGCCACCTCGCTCCGCTCGGAGGACGCT 1457  
QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
DB 1458 CGTGTGAGCCCTATGTGACATTTGCTGAGTTTACCGCTCTGGAGCGTGACCATGCG 1517  
QY 61 GluGlnSerValThrAlaPro 68  
DB 1518 GAGCAGAGCGTGGTGACAGCACCG 1541

RESULT 4  
US-09-144-178-5  
Sequence 5, Application US/09144178  
Patent No. 5989862  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihiro  
APPLICANT: NISHIDA, Eisuke  
TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/144,178  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/752,891  
FILING DATE: 20-NOV-1996  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1560 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1541  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 30..1541  
US-09-144-178-5

Alignment Scores:  
Pred. No.: 6.56e-36 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x US-09-144-178-5 (1-1560)

QY 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
DB 1338 CAAGACCCGACCTTAAACCTTCAGTCCACACACGACGACGAGCAGCTCCACG 1397  
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40  
DB 1398 TCTGACGAGGCGCTTCTCCGCTCCCGGCCGCCACCTCGCTCCGCTCGGAGGACGCT 1457  
QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60  
DB 1458 CGTGTGAGCCCTATGTGACATTTGCTGAGTTTACCGCTCTGGAGCGTGACCATGCG 1517  
QY 61 GluGlnSerValThrAlaPro 68  
DB 1518 GAGCAGAGCGTGGTGACAGCACCG 1541

RESULT 5  
US-09-406-854-1  
Sequence 1, Application US/09406854  
Patent No. 6140042  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihiro  
APPLICANT: NISHIDA, Eisuke  
TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR

TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1560 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1541  
NAME/KEY: mat\_peptide  
LOCATION: 30..1541  
US-08-752-891-1

Alignment Scores:  
Pred. No.: 6,566-36 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x US-08-752-891-1 (1-1560)

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSer 20  
Db 1338 CAAAGCCGACCTTAACCTTCAGTCCACACAGCAGCAGCAGCAGCAGCAGCAGCAGC 1397  
Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGlyAspGly 40  
Db 1398 TCTGACGAGGAGCCTCTCCGCTCCGCGCCGCTGCTCCGCTGAGGAGAGAGGT 1457  
Qy 41 ArgValGlnProTyrValAspPheAlaGluPheTyrArgLeuTyrPserValAspHisGly 60  
Db 1458 CGTGTGAGCCCTATGATGAGCTTTGCTGAGTTTACCGCCTCTGAGAGCGTGCATGGC 1517  
Qy 61 GlnGlnSerValValThrAlaPro 68  
Db 1518 GAGCAGAGCGTGTGACAGCACCG 1541

RESULT 2  
US-08-752-891-5  
Sequence 5, Application US/08752891  
Patent No. 5837819  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihito  
APPLICANT: NISHIDA, Eisuke  
TITLE OF INVENTION: TABI PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
City: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,891  
FILING DATE: 20-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-126282

FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1560 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1541  
NAME/KEY: mat\_peptide  
LOCATION: 30..1541  
US-08-752-891-5

Alignment Scores:  
Pred. No.: 6,566-36 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x US-08-752-891-5 (1-1560)

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSer 20  
Db 1338 CAAAGCCGACCTTAACCTTCAGTCCACACAGCAGCAGCAGCAGCAGCAGCAGCAGC 1397  
Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGlyAspGly 40  
Db 1398 TCTGACGAGGAGCCTCTCCGCTCCGCGCCGCTGCTCCGCTGAGGAGAGAGGT 1457  
Qy 41 ArgValGlnProTyrValAspPheAlaGluPheTyrArgLeuTyrPserValAspHisGly 60  
Db 1458 CGTGTGAGCCCTATGATGAGCTTTGCTGAGTTTACCGCCTCTGAGAGCGTGCATGGC 1517  
Qy 61 GlnGlnSerValValThrAlaPro 68  
Db 1518 GAGCAGAGCGTGTGACAGCACCG 1541

RESULT 3  
US-09-144-178-1  
Sequence 1, Application US/09144178  
Patent No. 5989862  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihito  
APPLICANT: NISHIDA, Eisuke  
TITLE OF INVENTION: TABI PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
City: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/144,178  
FILING DATE:

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 10, 2002, 05:36:06 ; Search time 68 Seconds

(without alignments)  
306.677 Million cell updates/sec

Title: US-09-830-144-4\_COPY\_437\_504

Perfect score: 359  
Sequence: 1 QSPFLTQSTNTHQTSSSS.....AEFYRLMSVDHGQSVWTAP 68

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US09830144/runat\_04122002\_141843\_5006/app\_query.fasta\_1.263  
-DB=Issued\_Patents.NA -QWMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blosum62 -FRANS=numa40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09830144 @CGN\_1\_1\_31 @runat\_04122002\_141843\_5006 -NCPV=6 -ICPV=3  
-NO XLPXY -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents.NA.\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	359	100.0	1560	2	US-08-752-891-1 Sequence 1, Appli
2	359	100.0	1560	2	US-08-752-891-5 Sequence 5, Appli
3	359	100.0	1560	2	US-09-144-178-1 Sequence 1, Appli
4	359	100.0	1560	2	US-09-144-178-5 Sequence 5, Appli
5	359	100.0	1560	3	US-09-406-854-1 Sequence 1, Appli
6	359	100.0	1560	3	US-09-406-854-5 Sequence 5, Appli
7	359	100.0	1560	4	US-09-529-279-1 Sequence 1, Appli
8	359	100.0	1568	4	US-09-529-279-42 Sequence 42, Appli
9	359	100.0	1569	4	US-09-529-279-10 Sequence 10, Appli
c 10	71.5	19.9	1108	5	PCT-US93-03035-1 Sequence 1, Appli
11	67.5	18.8	2559	2	US-09-070-060-2 Sequence 2, Appli
12	67.5	18.8	2559	3	US-09-357-746-2 Sequence 2, Appli

13	67.5	18.8	5359	2	US-09-070-060-1 Sequence 1, Appli
14	67.5	18.8	5372	3	US-09-357-746-1 Sequence 1, Appli
15	63.5	17.7	1101	4	US-09-152-060-23 Sequence 23, Appli
16	63.5	17.7	1173	3	US-08-706-216-5 Sequence 5, Appli
17	63.5	17.7	1473	4	US-09-152-060-43 Sequence 43, Appli
c 18	63.5	17.7	2784	4	US-08-857-076-39 Sequence 39, Appli
c 19	63.5	17.7	3017	4	US-08-857-076-52 Sequence 52, Appli
c 20	63.5	17.7	3119	4	US-08-857-076-53 Sequence 53, Appli
c 21	62	17.3	1395	2	US-08-553-367A-1 Sequence 1, Appli
c 22	62	17.3	1395	2	US-09-295-306-1 Sequence 1, Appli
c 23	62	17.3	1395	4	US-09-734-719-1 Sequence 1, Appli
c 24	62	17.3	5977	3	US-09-024-020B-1 Sequence 1, Appli
c 25	62	17.3	5977	3	US-09-425-043-1 Sequence 1, Appli
c 26	62	17.3	6007	3	US-09-024-020B-2 Sequence 2, Appli
c 27	62	17.3	6007	4	US-09-425-043-2 Sequence 2, Appli
c 28	62	17.3	6556	3	US-09-024-020B-7 Sequence 7, Appli
c 29	62	17.3	6556	3	US-09-425-043-7 Sequence 7, Appli
c 30	62	17.3	6586	3	US-09-024-020B-43 Sequence 43, Appli
c 31	62	17.3	6586	4	US-09-425-043-43 Sequence 43, Appli
c 32	62	17.3	6826	3	US-09-024-020B-8 Sequence 8, Appli
c 33	62	17.3	6826	4	US-09-425-043-8 Sequence 8, Appli
c 34	61.5	17.1	11303	4	US-08-961-527-115 Sequence 115, App
c 35	61.5	17.1	3478	1	US-08-530-492-1 Sequence 1, Appli
c 36	61.5	17.1	3478	4	US-08-906-517-1 Sequence 1, Appli
c 37	60.5	16.9	1518	2	US-08-997-080-88 Sequence 88, Appli
c 38	60.5	16.9	1518	2	US-08-997-362-88 Sequence 88, Appli
c 39	60.5	16.9	1518	3	US-08-873-970-88 Sequence 88, Appli
c 40	60.5	16.9	1518	4	US-09-095-855-88 Sequence 88, Appli
c 41	60.5	16.9	1518	4	US-09-324-542-88 Sequence 88, Appli
c 42	60.5	16.9	1518	4	US-09-205-426-88 Sequence 88, Appli
c 43	60.5	16.9	169998	4	US-09-676-610B-24 Sequence 24, Appli
c 44	60	16.7	903	5	PCT-US95-06406A-21 Sequence 21, Appli
c 45	60	16.7	1071	1	US-08-357-962-1 Sequence 1, Appli

#### ALIGNMENTS

#### RESULT 1

US-08-752-891-1  
; Sequence 1, Application US/08752891  
; Patent No. 5837819  
; GENERAL INFORMATION:  
; APPLICANT: MATSUMOTO, Kunihiko  
; APPLICANT: NISHIDA, Eisuke  
; TITLE OF INVENTION: TABI PROTEIN AND DNA CODING THEREFOR  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752,891  
; FILING DATE: 20-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-300856  
; FILING DATE: 28-OCT-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-126282  
; FILING DATE: 24-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER:  
; REFERENCE/DOCKET NUMBER: 17981/111  
; TELECOMMUNICATION INFORMATION:



```

Db 61 GGGGGCTTCCGGCTCCAGACCGGCTCACTTCCACCCGAGAGGATGCCGTGTG 120
|||
Qy 43 GluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGlyGluGln 62
|||
Db 121 GAGCCCTATGTGGACTTGTGCTGAGTTCTTACCGACTCTGGAGCGTGGACCGGGAGCAG 180
|||
Qy 63 SerValValThrAlaPro 68
|||
Db 181 AGCGTGATGACGGCACCT 198
|||

RESULT 15
BI181203/c
LOCUS      BI181203              777 bp      mRNA      linear      EST 10-JUL-2001
DEFINITION UNL-P-FN-ae-a-05-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-ae-a-05-0-UNL 3', mRNA sequence.
ACCESSION  BI181203
VERSION    BI181203.1 GI:14655612
KEYWORDS   EST.
SOURCE     pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
            1 (bases 1 to 777)
            Caetano, A.R., Johnson, R.K. and Pomp, D.
            Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
            Unpublished (2001)
            Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found, Not 1 site shown in beginning of sequence
            is likely internal to the message. The following repetitive
            elements were found in this cDNA sequence: 61-111,
            >GC rich#Low complexity
            Seq primer: M13 -29
            POLYA-No.

FEATURES             Location/Qualifiers
     source            1..777
                        /organism="Sus scrofa"
                        /strain="University of Nebraska, Lincoln Swine Selection
                        Lines"
                        /db_xref="taxon:9823"
                        /clone="UNL-P-FN-ae-a-05-0-UNL"
                        /clone_lib="UNL-P-FN"
                        /dev_stage="ADULT"
                        /lab_host="DH10B (Life Technologies)"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not 1; Site_2: Eco RI; The UNL-P-FN
                        library is a normalized library representing porcine
                        ovarian follicles, ranging between 2.0 to 10.0 mm in
                        diameter, collected during 7 days of the follicular phase
                        of the pig estrous cycle. This library was derived from
                        the library UNL-P-F2. The tag is a string of 5-6
                        nucleotides present between the Not 1 site and the
                        oligo-dT track. The library was constructed as described
                        by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
                        , 1996.
BASE COUNT      106 a      247 c      290 g      127 t      7 others
ORIGIN
Alignment Scores:
Pred. No.:      3,82e-30      Length:      777
Score:          345.00      Matches:      65
Percent Similarity: 95.59%      Conservative: 0
Best Local Similarity: 95.59%      Mismatches: 3
Query Match:    96.10%      Indels:      0
DB:             13      Gaps:      0

```

US-09-830-144-4\_COPY\_437\_504 (1-68) x BI181203 (1-777)

```

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20
|||
Db 317 CAGAGCCCGACCCCTGACCCCTGCAGTCCACCAACCCACCCAGCAGCAGTCCAGC 258
|||
Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40
|||
Db 257 TCCGACGGGGGCTCTTTTCGCTCCCGGCCACCCACTCGCTCCCGCCCGGCGGATGCG 198
|||
Qy 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60
|||
Db 197 CGCGTGGAGCCCTTACGTGGACTTCGACAGAGTTCTACCGCTGTGGAGCGTGGACCGGC 138
|||
Qy 61 GluGlnSerValValThrAlaPro 68
|||
Db 137 GAGCAGAGTGTGGCGACGGGCGCG 114
|||

```

Search completed: December 10, 2002, 07:26:07  
Job time : 2220 secs

DEFINITION UNL-P-FN-au-g-01-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone  
 ACCSSION UNL-P-FN-au-g-01-0-UNL 3', mRNA sequence.  
 VERSION B1181306  
 KEYWORDS B1181306.1 GI:14655715  
 SOURCE EST.  
 ORGANISM pig.  
 Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 REFERENCE 1 (bases 1 to 815)  
 AUTHORS Gaetano, A.R., Johnson, R.K. and Pomp, D.  
 TITLE Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Pomp, D  
 Department of Animal Science  
 University of Nebraska, Lincoln  
 Lincoln, NE 68583-0908, USA  
 Tel: 402 472 6416  
 Fax: 402 472 6362  
 Email: dpomp@unl.edu  
 Oligo-dt track not found, Not I site shown in beginning of sequence  
 is likely internal to the message. The following repetitive  
 elements were found in this cDNA sequence: 62-112,  
 >CC-rich#low\_complexity  
 Seq primer: M13 -29  
 POLYA=NO.

FEATURES  
 source  
 1..815  
 Location/Qualifiers  
 /organism="Sus scrofa"  
 /strain="University of Nebraska, Lincoln Swine Selection  
 Lines"  
 /db\_xref="taxon:9823"  
 /clone="UNL-P-FN-au-g-01-0-UNL"  
 /clone\_1ib="UNL-P-FN"  
 /dev\_stage="ADULT"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pMT73D-Pac (Pharmacia) with a modified  
 polylinker, Site\_1: Not I; Site\_2: Eco RI; The UNL-P-FN  
 library is a normalized library representing porcine  
 ovarian follicles, ranging between 2.0 to 10.0 mm in  
 diameter, collected during 7 days of the follicular phase  
 of the pig estrous cycle. This library was derived from  
 the library UNL-P-F2. The tag is a string of 5-6  
 nucleotides present between the Not I site and the  
 oligo-dt track. The library was constructed as described  
 by Bonaldo, Lennon and Soares, Genome Research 6: 791-805  
 , 1996.  
 TAG\_SEQ=None found"  
 BASE COUNT 109 a 266 c 306 g 134 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 8..14e-31 Length: 815  
 Score: 351.00 Matches: 66  
 Percent Similarity: 97.06% Conservative: 0  
 Best Local Similarity: 97.06% Mismatches: 2  
 Query Match: 97.77% Indels: 0  
 DB: 13 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x B1181306 (1-815)  
 QY 1 GlnserProthleuThrluGlnserThraThrhIsThnGlnserSerSerSer 20  
 Db 318 CAGAGCCCGAGCCCTGACCTGACGACCAACACCCACAGAGAGACTTCAGC 259  
 QY 21 SeraspGlyglyleuPheAgtSerAArgProAlahIsSerleuProProGlygluAspGly 40  
 Db 258 TCCACACGGGGAGCCCTCTTCTCCCTCCGGCCACCACTGGCTCCGGCCGGAGAGATGGC 199  
 QY 41 ArgValGluProGlyValAAsPheAlaGluPheTYrArgLeuTrpSerValAspHiGly 60  
 Db 198 CGCCTGAGGCGCTACGAGGACTTCGCGAGATTCTACCGCCTGTGGAGGTGACACACGCG 139

QY 61 GluGlnSerValValThrAlaPro 68  
 Db 138 GAGCAGAGTGTGTGGCAGCGCGCGC 115

RESULT 14  
 LOCUS BM934363  
 DEFINITION 644 bp mRNA linear EST 29-APR-2002  
 UI-M-CGOp-b11-a-03-0-UI.r1 NIH BMAP Ret4.S2 Mus musculus cDNA clone  
 ACCSSION UI-M-CGOp-b11-a-03-0-UI 5', mRNA sequence.  
 VERSION BM934363  
 KEYWORDS BM934363.1 GI:19393515  
 SOURCE EST.  
 ORGANISM house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 644)  
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: MEST@mail.nih.gov  
 Tissue Procurement: Dr. Xin-Yuan Fu, Yale University School of  
 Medicine  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 REVERSE.

FEATURES  
 source  
 1..644  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-CGOp-b11-a-03-0-UI"  
 /clone\_1ib="NIH BMAP Ret4.S2"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pMT73D-Pac (Pharmacia) with a modified  
 polylinker, Site\_1: Not I; Site\_2: Eco RI; The  
 NIH\_BMAP\_Ret4.S2 library is a subtracted library,  
 ultimately derived from mouse retina tissue libraries at  
 various stages of development. For a detailed description  
 of the library from which this clone was derived, please  
 visit our web site at brainest.eng.uiowa.edu. The tissue  
 for this library was contributed by Dr. Xin-Yuan Fu, Yale  
 University School of Medicine"

BASE COUNT 138 a 210 c 174 g 121 t 1 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.75e-30 Length: 644  
 Score: 347.00 Matches: 65  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 98.48% Mismatches: 0  
 Query Match: 96.66% Indels: 0  
 DB: 14 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x BM934363 (1-644)  
 QY 3 ProthleuThrluGlnserThraThrhIsThnGlnserSerSerSer 22  
 Db 1 CCACCTCAGCCCTGAGTCACCAACACGACACCAAGAGAGAGCTCAGCTGAC 60  
 QY 23 GlyGlyleuPheAgtSerAArgProAlahIsSerleuProProGlygluAspGly 42

**MEDLINE**  
**COMMENT**  
 97044477  
 Contact: Tuggle CK  
 Molecular Genetics Laboratory, Department of Animal Science  
 Iowa State University  
 201 Kildee Hall, Ames, IA 50011-3150, USA  
 Tel: 5152944252  
 Fax: 5152942401  
 Email: cktuggle@iastate.edu  
 Oligo-dT track not found, Not I site shown in beginning of sequence  
 is likely internal to the message. cDNA Library Preparation: M.B.  
 Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab,  
 University of Iowa Clone distribution: clones will be available  
 through Research Genetics (www.resgen.com) The following repetitive  
 elements were found in this cDNA sequence: 60-110,  
 >GC rich#Low complexity  
 Seq primer: M13 Forward  
 POLYA=No.

**FEATURES**  
 source  
 1..599  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone\_lib="MI-P-CP0-nvn-g-07-0-UI"  
 /clone\_lib="MI-P-CP0"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: ScaRI; The MI-P-CP0  
 library is derived from uterus. For a detailed description  
 of the library from which this clone was derived, please  
 visit our web site at http://pigest.genome.iastate.edu/.  
 The procedure used to create this library has been  
 previously described (Bonaldo, Lennon and Soares, Genome  
 Research 6:791-806, 1996)  
 TAG\_SEQ=None found"  
 77 a \_193 c 238 g 91 t

**BASE COUNT**  
 77 a \_193 c 238 g 91 t

**ORIGIN**

Alignment Scores:  
 Pred. No.: 5,46e-31 Length: 599  
 Score: 351.00 Matches: 66  
 Percent Similarity: 97.06% Conservative: 0  
 Best Local Similarity: 97.06% Mismatches: 2  
 Query Match: 97.77% Indels: 0  
 DB: 13 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x BI401530 (1-599)

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
 |||||  
 Db 316 CAGAGCCGACCCCTGACCTGCAGTCCACCAACCCACCCAGCAGCAGTCCAGC 257  
 |||||  
 Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40  
 |||||  
 Db 256 TCCGACGGGGGGCTCTTTTCCTCCGGCCACCCACTCGCTCCGCGCGGGAGATGGC 197  
 |||||  
 Qy 41 ArgValGluProTyValAspPheAlaGluPheTyArgLeuTrpSerValAspHisGly 60  
 |||||  
 Db 196 CGCGTGAGGCGCCCTACGTGGAGTTCGCGGAGTTCCTACCGCCTGTGGAGCGTGGACCGGC 137  
 |||||  
 Qy 61 GluGlnSerValValThrAlaPro 68  
 |||||  
 Db 136 GAGCAGAGTGTGGCAGCGCGCGC 113  
 |||||

**RESULT 12**  
 BI184403/c  
 LOCUS  
 DEFINITION  
 UNL-P-FN-cn-c-04-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone  
 UNL-P-FN-cn-c-04-0-UNL 3', mRNA sequence.  
 BI184403  
 ACCESSION  
 BI184403.1 GI:14658812  
 VERSION  
 KEYWORDS  
 EST.  
 SOURCE  
 Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 1 (bases 1 to 648)  
 Caetano,A.R., Johnson,R.K. and Pomp,D.  
 Generation and sequence characterization of a normalized cDNA  
 library from swine ovarian follicles  
 Unpublished (2001)  
 Contact: Pomp, D  
 Department of Animal Science  
 University of Nebraska, Lincoln  
 Lincoln, NE 68583-0908, USA  
 Tel: 402 472 6416  
 Fax: 402 472 6362  
 Email: dpomp@unl.edu  
 Oligo-dT track not found, Not I site shown in beginning of sequence  
 is likely internal to the message. The following repetitive  
 elements were found in this cDNA sequence: 61-111,  
 >GC rich#Low complexity  
 Seq primer: M13 -29  
 POLYA=No.

**FEATURES**  
 Location/Qualifiers  
 1..648  
 /organism="Sus scrofa"  
 /strain="University of Nebraska, Lincoln Swine Selection  
 Lines"  
 /db\_xref="taxon:9823"  
 /clone\_lib="UNL-P-FN-cn-c-04-0-UNL"  
 /dev\_stage="ADULT"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; The UNL-P-FN  
 library is a normalized library representing porcine  
 ovarian follicles, ranging between 2.0 to 10.0 mm in  
 diameter, collected during 7 days of the follicular phase  
 of the pig estrous cycle. This library was derived from  
 the library UNL-P-FN. The tag is a string of 5-6  
 nucleotides present between the Not I site and the  
 oligo-dT track. The library was constructed as described  
 by Bonaldo, Lennon and Soares, Genome Research 6: 791-806  
 , 1996.  
 TAG\_SEQ=None found"  
 85 a \_205 c 252 g 99 t 7 others

**BASE COUNT**  
 85 a \_205 c 252 g 99 t 7 others

**ORIGIN**

Alignment Scores:  
 Pred. No.: 6,05e-31 Length: 648  
 Score: 351.00 Matches: 66  
 Percent Similarity: 97.06% Conservative: 0  
 Best Local Similarity: 97.06% Mismatches: 2  
 Query Match: 97.77% Indels: 0  
 DB: 13 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x BI184403 (1-648)

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
 |||||  
 Db 317 CAGAGCCGACCCCTGACCTGCAGTCCACCAACCCACCCAGCAGCAGTCCAGC 258  
 |||||  
 Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40  
 |||||  
 Db 257 TCCGACGGGGGGCTCTTTTCCTCCGGCCACCCACTCGCTCCGCGCGGGAGATGGC 198  
 |||||  
 Qy 41 ArgValGluProTyValAspPheAlaGluPheTyArgLeuTrpSerValAspHisGly 60  
 |||||  
 Db 197 CGCGTGAGGCGCCCTACGTGGAGTTCGCGGAGTTCCTACCGCCTGTGGAGCGTGGACCGGC 138  
 |||||  
 Qy 61 GluGlnSerValValThrAlaPro 68  
 |||||  
 Db 137 GAGCAGAGTGTGGCAGCGCGCGC 114  
 |||||

**RESULT 13**  
 BI181306/c  
 LOCUS  
 BI181306 815 bp mRNA linear EST 10-JUL-2001



Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab, University of Iowa Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 62-112,  
>GC rich#low complexity  
Seq primer: M13 Forward  
POLYA=No.

# FEATURES

## source

1..527  
/organism="Sus scrofa"  
/strain="crossbred"  
/db\_xref="taxon:9823"  
/clone="MI-P-AV1-ncg-a-03-0-UI"  
/lab\_host="MDH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: EcoRI; The MI-P-AV1 library is normalized library derived from the MI-P-AV0 library, ultimately derived from placenta tissue. For a detailed description of the library from which this clone was derived, please visit our web site at <http://bigest.genome.iastate.edu/>. The procedure used to create this library has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)  
TAG\_SEQ=None found"

BASE COUNT 67 a 166 c 216 g 78 t  
ORIGIN

## Alignment Scores:

Pred. No.: 4,63e-31 Length: 527  
Score: 351.00 Matches: 66  
Percent Similarity: 97.06% Conservative: 0  
Best Local Similarity: 97.06% Mismatches: 2  
Query Match: 97.77% Indels: 0  
Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x B1400511 (1-527)

QY 1 GlnSerProThleuThleuGlnSerThraAnthrhIsthGlnSerSerSerSer 20  
DB 318 CAGAGCCCAACCTGACCTGACGACCAACCAACCAACCAACCAACCAACCTCCAC 259  
QY 21 SerASPGLVGLyleuPheArgSerArgProAlaHisSerleuProproGlyGluASPGLY 40  
DB 258 TCCGAGCGGGGGCTCTTTCGCTCCCGGCCCAACCACTCCGCCCGCGGAGATGGC 199  
QY 41 ArgValGluProTyValAspPheAlaGluPheTyArgleuTpsSerValAspHisGly 60  
DB 198 CGCGTGAGACCTACGTGAGACTTCGCGAGTTCTACCGCTGTGAGCGTGACCAAGC 139  
QY 61 GluGlnSerValValThrAlaPro 68  
DB 138 GAGCAGAGTGTGGCAGCGGCCG 115

RESULT 10 582 bp mRNA linear EST 17-SEP-2001  
B1681846  
LOCUS B1681846 461334 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
DEFINITION B1681846  
ACCESSION B1681846.1 GI:15634771  
VERSION  
KEYWORDS EST.  
SOURCE  
ORGANISM Bos taurus  
Bovidae; Bovinae; Bos.

## REFERENCE

1 (bases 1 to 582)  
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Cass,E., Wray,J.E., White,J., Cho,J., Fahrrenkrug,S.C., Bennett,  
G.U., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,  
Perteira,G., Holt,I., Karanymcheva,S., Liang,F., Quackenbush,J. and  
Keefe,J.W.  
Sequence evaluation of four pooled-tissue normalized bovine cDNA

JOURNAL  
MEDLINE  
COMMENT  
libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)  
21180013  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smithemil.marc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mimatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTGACGACG  
Plate: 135 row: 0 column: 15  
Seq primer: ATTAGGTGACACTATAG.

# FEATURES

## source

1..582  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 1BOV"  
/tissue\_type="pooled"  
/lab\_host="MDH10B"  
/note="Vector: pCMV SPOR6; Site 1: Not1, Site 2: SalI,  
Library made from pooled tissue from lymph node, ovary,  
fat, hypothalamus, and pituitary."

BASE COUNT 93 a 212 c 179 g 98 t  
ORIGIN

## Alignment Scores:

Pred. No.: 5,26e-31 Length: 582  
Score: 351.00 Matches: 67  
Percent Similarity: 98.53% Conservative: 0  
Best Local Similarity: 98.53% Mismatches: 1  
Query Match: 97.77% Indels: 0  
Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x B1681846 (1-582)

QY 1 GlnSerProThleuThleuGlnSerThraAnthrhIsthGlnSerSerSerSer 20  
DB 80 CAGAGCCCAACCTGACCTGACGACCAACCAACCAACCAACCAACCAACCTCCAC 139  
QY 21 SerASPGLVGLyleuPheArgSerArgProAlaHisSerleuProproGlyGluASPGLY 40  
DB 140 TCCGAGCGGGGGCTTTCGCTCCCGGCCCAACCACTCCGCCCGCTGAGATGGC 199  
QY 41 ArgValGluProTyValAspPheAlaGluPheTyArgleuTpsSerValAspHisGly 60  
DB 200 CGCGTGAGACCTACGTGAGACTTCGCGAGTTCTACCGCTGTGAGCGTGACCAATGCC 259  
QY 61 GluGlnSerValValThrAlaPro 68  
DB 260 GAGCAGAGTGTGGCAGCGGCCG 283

RESULT 11 599 bp mRNA linear EST 14-AUG-2001  
B1401530  
LOCUS B1401530/c 599 bp mRNA linear EST 14-AUG-2001  
DEFINITION B1401530  
ACCESSION B1401530  
VERSION B1401530.1 GI:15180591  
KEYWORDS EST.  
SOURCE  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euteria; Cetartiodactyla; Suidae; Suidae; Sus.

## REFERENCE

1 (bases 1 to 599)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)

```

FEATURES
  source
    POLYA=No.
    Location/Qualifiers
      1..334
        /organism="Sus scrofa"
        /strain="crossbreed"
        /db_xref="taxon:9823"
        /clones="MI-P-Ay1-ngr-c-02-0-UI"
        /clone_lib="MI-P-Ay1"
        /lab_host="DH10B (Life Technologies)"
        /note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: EcoRI; The MI-P-Ay1
library is normalized library derived from the MI-P-Ay1
library, ultimately derived from placenta tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
http://pigest.genome.tastate.edu/. The procedure used to
create this library has been previously described (Bonaldo
, Lennon and Soares, Genome Research 6: 791-806, 1996)
TAG_SEQ=None found"
      BASE COUNT 43 a _111 c 136 g 44 t
      ORIGIN

Alignment Scores:
  Pred. No.: 2,56e-31 Length: 334
  Score: 351.00 Matches: 66
  Percent Similarity: 97.06% Conservative: 0
  Best Local Similarity: 97.06% Mismatches: 2
  Query Match: 97.77% Indels: 0
  DB: 12 Gaps: 0

US-09-830-144-4_COPY_437_504 (1-68) x BF710376 (1-334)
QY 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20
Db 317 CAGAGCCCGACCGCTGACCTTCAGCTCCACCACACCCACACCCAGAGCGAGTCCAGC 258
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40
Db 257 TCCGACGGGGGGCTCTTCGCTCCCGGCCACCCACCTCGCTCCCGCGCGGAGATGCG 198
QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60
Db 197 CGCGTGGAGCCCTAGTGGACTTCGCGGAGTTCACCGCTCTGGAGCGTGGACACCGC 138
QY 61 GluGlnSerValValThrAlaPro 68
Db 137 GAGCAGAGTGTGGCGACGGCGCG 114

RESULT 8
LOCUS BG382061 378 bp mRNA linear EST 12-MAR-2001
DEFINITION 297803 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BG382061
VERSION BG382061.1 GI:13306533
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred

FEATURES
  source
    v0.980904.e. Vector identified by cross_match with the -minscore 18
    and -minmatch 12 options.
    PCR Primers
      FORWARD: AGGAACAGCTATGACCAT
      BACKWARD: GTTTCCTCCAGTCAGCAGC
    Plate: 2 row: 0 column: 16
    Seq primer: ATTTAGGTGACACTAG.
    Location/Qualifiers
      1..378
        /organism="Sus scrofa"
        /db_xref="taxon:9803"
        /clone_lib="MARC 1P1G"
        /tissue_type="pooled"
        /lab_host="DH10B"
        /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
      BASE COUNT 52 a 154 c 123 g 49 t
      ORIGIN

Alignment Scores:
  Pred. No.: 3,01e-31 Length: 378
  Score: 351.00 Matches: 66
  Percent Similarity: 97.06% Conservative: 0
  Best Local Similarity: 97.06% Mismatches: 2
  Query Match: 97.77% Indels: 0
  DB: 12 Gaps: 0

US-09-830-144-4_COPY_437_504 (1-68) x BG382061 (1-378)
QY 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20
Db 71 CAGAGCCCGACCGCTGACCTTCAGCTCCACCACACCCACACCCAGAGCGAGTCCAGC 130
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40
Db 131 TCCGACGGGGGGCTCTTCGCTCCCGGCCACCCACCTCGCTCCCGCGCGGAGATGCG 190
QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60
Db 191 CGCGTGGAGCCCTAGTGGACTTCGCGGAGTTCACCGCTCTGGAGCGTGGACACCGC 250
QY 61 GluGlnSerValValThrAlaPro 68
Db 251 GAGCAGAGTGTGGCGACGGCGCG 274

RESULT 9
LOCUS BI400511/c 527 bp mRNA linear EST 14-AUG-2001
DEFINITION MI-P-Ay1-ngr-a-03-0-UI.s1 MI-P-Ay1 Sus scrofa cDNA clone
ACCESSION BI400511
VERSION BI400511.1 GI:15179572
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
COMMENT Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu
Oligo-dT track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.

```



Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Cloned through the I.M.A.G.E. Consortium/LLNL at:  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM765 row: e column: 05  
 High quality sequence stop: 583.  
 Location/Qualifiers  
 1. .705  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3593452"  
 /clone\_lib="NCI CGAP Maml"  
 /tissue\_type="tumor, biopsy sample"  
 /dev\_stage="10 months, virgin"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"

BASE COUNT 159 a 231 c 194 g 121 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,77e-31 Length: 705  
 Score: 356.00 Matches: 67  
 Percent Similarity: 100.00% Conservatives: 1  
 Best Local Similarity: 98.53% Mismatches: 0  
 Query Match: 99.16% Indels: 0  
 DB: 10 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x BE375985 (1-705)

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
 Db 132 CAGAGCCCCACTGTGACCTTGCAGTCCACCAACAGCACGCCAGCAGCAGCTCCAGC 191  
 Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGly 40  
 Db 192 TCTGACGGGGGCTCTTCGCTCCAGCCGGCTCACTCACTCCACCGGAGAGATGC 251  
 Qy 41 ArgValGluProTrpValAspPheAlaGluPheTyArgLeuTrpSerValAspHisGly 60  
 Db 252 CGTGTGAGGCCCTATGTGGACTTGTGTGAGTTCTACCGACTCTGAGCGTGGACCGC 311  
 Qy 61 GluGlnSerValThrAlaPro 68  
 Db 312 GAGCAGAGCGTGATGACGGCACT 335

RESULT 4  
 B1253992  
 LOCUS B1253992 782 bp mRNA linear EST 17-JUL-2001  
 DEFINITION 602975130F1 NIH\_MGC\_12 Homo sapiens CDNA clone IMAGE:5114622 5',  
 mRNA sequence.  
 ACCESSION B1253992  
 VERSION B1253992.1 GI:14805965  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 782)  
 NIH-MGC http://img.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Cloned through the I.M.A.G.E. Consortium/LLNL at:  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM1278 row: k column: 07  
 High quality sequence stop: 657.  
 Location/Qualifiers  
 1. .782  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5114622"  
 /clone\_lib="NIH\_MGC\_12"  
 /tissue\_type="cervical carcinoma cell line"  
 /lab\_host="DH10B"  
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.4 kb. Library prepared by Life  
 Technologies."

BASE COUNT 162 a 288 c 214 g 118 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2,64e-31 Length: 782  
 Score: 355.00 Matches: 67  
 Percent Similarity: 98.53% Conservatives: 0  
 Best Local Similarity: 98.53% Mismatches: 1  
 Query Match: 98.89% Indels: 0  
 DB: 13 Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x B1253992 (1-782)

Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20  
 Db 306 CAAAGCCCGACCTTAACCTCGAGTCCACCAACAGCACGCCAGCAGCAGCTCCAGC 365  
 Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGly 40  
 Db 366 TCTGACGGAGGGCTCTTCGCTCCCGGCCCTGCCACTCGCTCCCGCTCGCGAGGACGT 425  
 Qy 41 ArgValGluProTrpValAspPheAlaGluPheTyArgLeuTrpSerValAspHisGly 60  
 Db 426 CGTGTGAGCCCTATGTGACTTTGCTGAGTTTACCGCCTCTGGAGCGTGGACCATGSC 485  
 Qy 61 GluGlnSerValThrAlaPro 68  
 Db 486 GAGCAGAGCGTGGTGACACCG 509

RESULT 5  
 BQ925425  
 LOCUS BQ925425 1385 bp mRNA linear EST 20-AUG-2002  
 DEFINITION AGENCOURT\_8801162 NIH\_MGC\_129 Mus musculus CDNA clone IMAGE:6310974  
 5', mRNA sequence.  
 ACCESSION BQ925425  
 VERSION BQ925425.1 GI:22340456  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 1385)  
 NIH-MGC http://img.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: Susan L. Sullivan, PhD.  
 CDNA Library Preparation: ResGen, Invitrogen Corp  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Cloned through the I.M.A.G.E. Consortium/LLNL at:  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

JOURNAL  
COMMENT  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: DCM/DTP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: image.lnl.gov  
Place: LNC820 row: k column: 23  
High quality sequence stop: 652.

FEATURES  
source  
1. 974  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="N1H\_MGC\_9"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: ovary; Vector: pOT7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 209 a 311 c 278 g 176 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.2e-31 Length: 974  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) X BE898567 (1-974)

QY 1 GlnSerProThLeuThleuGlnSerThraThrHisThrGlnSerSerSerSer 20  
Db 433 CAAAGCCCGACCTTAACCTCGAGTCACCAACGACGACGAGACGACTCCAGC 492

QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGlyAspGly 40  
Db 493 TCTGACGGAGGCTCTTCCTCCGCGCCGCGCACTCGCTCCGCGGAGAGCGGT 552

QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTyrPserValAspHisGly 60  
Db 553 CGGTGTGAGCCCTATGTGAGCTTGTGAGTTTACCGCTCTGAGCGTGAGACCATGGC 612

QY 61 GluGlnSerValValThrAlaPro 68  
Db 613 GAGCAGAGCGTGTGACAGCAGCG 636

RESULT 2  
LOCUS AM258218 496 bp mRNA linear EST 03-APR-2000  
DEFINITION ug31h07.y1 NCI CGAP MamM Mus musculus cDNA clone IMAGE:2811037 5' similar to TR:Q15750 Q15750 TAKI BINDING PROTEIN.; mRNA sequence.  
ACCESSION AM258218  
VERSION AM258218.1 GI:6631199  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 496)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)

COMMENT  
Other ESTs: ug31h07.x1  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: www-bio.lnl.gov/bdip/image/image.html  
Seq primer: -40RP from Gibco  
High quality sequence stop: 439.

FEATURES  
source  
1. 496  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_id="IMAGE:2811037"  
/clone\_id="NCI\_CGAP\_MamM"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="7 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NciI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"

BASE COUNT 113 a 158 c 140 g 85 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.12e-31 Length: 496  
Score: 356.00 Matches: 67  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.53% Mismatches: 0  
Query Match: 99.16% Indels: 0  
DB: Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) X AM258218 (1-496)

QY 1 GlnSerProThLeuThleuGlnSerThraThrHisThrGlnSerSerSerSer 20  
Db 17 CAAAGCCCGACCTTAACCTCGAGTCACCAACGACGACGAGACGACTCCAGC 76

QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGlyAspGly 40  
Db 77 TCTGACGGAGGCTCTTCCTCCGCGCCGCGCACTCGCTCCGCGGAGAGTGGC 136

QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTyrPserValAspHisGly 60  
Db 137 CGGTGTGAGCCCTATGTGAGCTTGTGAGTTTACCGACTCTGAGCGTGAGCAGCGGC 196

QY 61 GluGlnSerValValThrAlaPro 68  
Db 197 GAGCAGAGCGTGTGACAGCAGCG 220

RESULT 3  
LOCUS BE375985 705 bp mRNA linear EST 21-JUL-2000  
DEFINITION 601299419P1 NCI\_CGAP MamM Mus musculus cDNA clone IMAGE:3593452 5', mRNA sequence.  
ACCESSION BE375985  
VERSION BE375985.1 GI:9321350  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 705)  
AUTHORS NIH-MGC http://mgc.ncbi.nlm.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

QM protein - nucleic search, using frame\_plus\_p2n model  
Run on: December 10, 2002, 05:34:46 ; Search time 2113 Seconds  
(without alignments)  
497.847 Million cell updates/sec  
Title: US-09-830-144-4\_COPY\_437\_504  
Perfect score: 359  
Sequence: 1 OSPTLTLOSTNTHQTSSSSS.....AEFYRLWSVDHGQSVWTAP 68

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV=xlp  
-MODEL=frame+ p2n model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US09830144/runat\_04122002\_141843\_4986/app\_query.fasta\_1.263  
-DB=EST -QPMI=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09830144 @CGN\_1\_1\_1716 @runat\_04122002\_141843\_4986 -NCPU=6 -ICPU=3  
-NO XLPXY -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_nam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_othr:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	359	100.0	974	12	BE988567	BE988567 601681494
2	356	99.2	496	10	AW258218	AW258218 ug31h07.y
3	356	99.2	705	10	BE375985	BE375985 601229419
4	355	98.9	782	13	BI253992	BI253992 602975130
5	355	98.9	1385	14	BQ925425	BQ925425 AGENCOURT
6	353	98.3	997	12	BG115732	BG115732 602317018
C 7	351	97.8	334	12	BF710376	BF710376 MI-P-AV1-
C 8	351	97.8	378	12	EG382061	EG382061 297803 MA
C 9	351	97.8	527	13	BI400511	BI400511 MI-P-AV1-
C 10	351	97.8	582	13	BI681846	BI681846 461334 MA
C 11	351	97.8	599	13	BI401530	BI401530 MI-P-CFO-
C 12	351	97.8	648	13	BI184403	BI184403 UNL-P-FN-
C 13	351	97.8	815	13	BI181306	BI181306 UNL-P-FN-
C 14	347	96.7	644	14	BM934363	BM934363 UI-M-CGOP
C 15	345	96.1	777	13	BI181203	BI181203 UNL-P-FN-
C 16	327	91.1	919	12	BF163305	BF163305 601771849
C 17	322	89.7	337	12	BF712308	BF712308 MI-P-A3-a
C 18	311	86.6	325	10	AW658281	AW658281 94183 MAR
C 19	309	86.1	617	12	EG710962	EG710962 pglin.pk0
C 20	287	79.9	1107	12	BE902307	BE902307 601676585
C 21	275	76.6	1576	11	AK009321	AK009321 Mus muscu
C 22	267	74.4	450	9	AA674170	AA674170 vp97a05.r
C 23	252	70.2	584	12	BE901639	BE901639 601677788
C 24	251.5	70.1	610	14	BQ387712	BQ387712 NISC.mn25
C 25	245.5	68.4	818	12	BE746542	BE746542 601580107
C 26	171.5	47.8	617	13	BI981448	BI981448 fu53e22.x
C 27	170	47.4	417	10	BE167965	BE167965 CM2-HT051
C 28	170	47.4	614	13	EG913752	EG913752 602810906
C 29	159.5	44.4	1016	17	CNS021V1	AL203302 Tetraodon
C 30	135	37.6	867	14	BQ431917	BQ431917 AGENCOURT
C 31	105	29.2	1102	13	BM9303BR	AL255744 Tetraodon
C 32	99	27.6	267	13	CNS03700	BM287700 528490 MA
C 33	97	27.0	501	9	AL602977	AL602977 DKF7p6661
C 34	84.5	23.5	546	9	AA707854	AA707854 zh24g12.s
C 35	83	23.1	481	9	AA198507	AA198507 mu19e08.r
C 36	78	21.7	300	9	AA403881	AA403881 vd80a02.r
C 37	78	21.7	863	17	CNS04AQ7	AL282184 Tetraodon
C 38	78	21.7	904	17	CNS029U9	AL187722 Tetraodon
C 39	78	21.7	925	17	CNS022SP	AL178594 Tetraodon
C 40	77	21.4	435	9	AL818083	AL818083 AL818083
C 41	77	21.4	457	12	BG262490	BG262490 WHE0936_F
C 42	77	21.4	494	12	BG262699	BG262699 WHE0940_C
C 43	76	21.2	563	17	AQ060318	AQ060318 CIT-HSP-2
C 44	75	20.9	461	9	AA218321	AA218321 mv74g07.r
C 45	75	20.9	633	14	BU009263	BU009263 QGH9P14.y

ALIGNMENTS

RESULT 1  
BE988567  
LOCUS BE988567 974 bp mRNA linear EST 29-SEP-2000  
DEFINITION 601681494F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3951886 5', mRNA sequence.  
ACCESSION BE988567  
VERSION BE988567.1 GI:10365176  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1. (bases 1 to 974)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)







```
Db 52 CAGACTCCGGCGCTGACAGCTCCACAGCTTACAGGTAGGGAGCTGGCCGAGGGCTGGCCAG 111
Qy 15 -----GlnSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 30
Db 112 GGGCGGCTGCTGCTCCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 171
Qy 31 AlaHisSerLeuProGlyGlyValAspGly 40
Db 172 CATCACCACTTGCTCTCCGCGGAGGGAAGGA 201

RESULT 11
US-09-764-864-1682/c
; Sequence 1682, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIORITY FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1682
; LENGTH: 8918
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-1682

Alignment Scores:
Pred. No.: 93.2 Length: 8918
Score: 68.00 Matches: 19
Percent Similarity: 47.06% Conservative: 5
Best Local Similarity: 37.25% Mismatches: 13
Query Match: 18.94% Indels: 14
DB: 2 Gaps: 2

US-09-830-144-4_COPY_437_504 (1-68) x US-09-764-864-1682 (1-8918)

Qy 23 GlyGlyLeuPheArgSer--ArgProAlaHis----- 32
Db 7922 GGAGGATCTTCAGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 7863

Qy 33 -----SerLeuProGlyGlyValAspGlyArgValGluProTyrValAspPhe 48
Db 7862 TAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 7803
Qy 49 AlaGluPheTyrArgLeuTyrSerValAspHis 59
Db 7802 GCGGGTTTCGCTCTCTTCATAGTATGATCAT 7770

RESULT 12
US-09-764-864-1683/c
; Sequence 1683, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIORITY FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1683
; LENGTH: 8919
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-1683

Alignment Scores:
Pred. No.: 93.2 Length: 8919
Score: 68.00 Matches: 19
Percent Similarity: 47.06% Conservative: 5
Best Local Similarity: 37.25% Mismatches: 13
Query Match: 18.94% Indels: 14
DB: 2 Gaps: 2
```

```
Percent Similarity: 47.06% Conservative: 5
Best Local Similarity: 37.25% Mismatches: 13
Query Match: 18.94% Indels: 14
DB: 10 Gaps: 2

US-09-830-144-4_COPY_437_504 (1-68) x US-09-764-864-1683 (1-8919)

Qy 23 GlyGlyLeuPheArgSer--ArgProAlaHis----- 32
Db 7923 GGAGGATCTTCAGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 7864

Qy 33 -----SerLeuProGlyGlyValAspGlyArgValGluProTyrValAspPhe 48
Db 7863 TAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 7804
Qy 49 AlaGluPheTyrArgLeuTyrSerValAspHis 59
Db 7803 GCGGGTTTCGCTCTCTTCATAGTATGATCAT 7771

RESULT 13
US-09-938-842A-2555/c
; Sequence 2555, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIORITY FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2555
; LENGTH: 966
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2555

Alignment Scores:
Pred. No.: 15.8 Length: 966
Score: 64.50 Matches: 19
Percent Similarity: 53.33% Conservative: 13
Best Local Similarity: 31.67% Mismatches: 19
Query Match: 17.97% Indels: 9
DB: 2 Gaps: 2

US-09-830-144-4_COPY_437_504 (1-68) x US-09-938-842A-2555 (1-966)

Qy 4 ThrieuThrieuGlnSerThraen-----ThrieuThrieuGlnSerSerSerSer 21
Db 865 ACTCTCTCTTACTTCTTCAAAATGATCTTGTGACGCGTCAAAATGCGCTTGTGACGCGTCA 806

Qy 22 AspGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGlyValAspPhe 39
Db 805 AATCTCATCTGTGTTCCAAACCTCAATGATCTTACACCTTGTGCAAGTGTCTT 746
Qy 40 -----GlyArgValGluProTyrValAspPheAlaGluPheTyrArgLeu 54
Db 745 CAGAGTATCATACAAAGATTGCCACTTCAATAGACTTCAGATGACATCTTCCTG 686

RESULT 14
US-10-027-806-2
; Sequence 2, Application US/10027806
; Patent No. US20020160476A1
; GENERAL INFORMATION:
```



```
|||||
Db 1375 TCTGACGAGGCGCTCTTCCGCTCCCGGCGCCGACCTCGCTCCCGCTGCGAGGAGCGGT 1434
QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTyrPseValAspHisGly 60
Db 1435 CGTGTGAGCCCTATGATGAGACTTGTGAGATTTCACGCCCTCTGAGACGTCGACATGCG 1494
QY 61 GluGlnSerValValThrAlaPro 68
Db 1495 GAGCAGAGCGGTGTGACAGCACCG 1518

RESULT 6
US-09-764-877-3349
; Sequence 3349, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3349
; LENGTH: 16877
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3349

Alignment Scores:
Pred. No.: 1,82e-35 Length: 16877
Score: 359.00 Matches: 68
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0

US-09-830-144-4_COPY_437_504 (1-68) x US-09-764-877-3349 (1-16877)
QY 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSer 20
Db 15011 CAAAGCCCGACCTTAACCTGACGTCACCAACAGCAGCAGCAGCAGCAGCAGCAGCAGC 15070
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGly 40
Db 15071 TCTGACGAGGCGCTCTTCCGCTCCGCGCCGCCCACTCGCTCCCGCTGCGAGGAGCGGT 15130
QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTyrPseValAspHisGly 60
Db 15131 CGTGTGAGCCCTATGATGAGACTTGTGAGATTTCACGCCCTCTGAGACGTCGACATGCG 15190
QY 61 GluGlnSerValValThrAlaPro 68
Db 15191 GAGCAGAGCGGTGTGACAGCACCG 15214

RESULT 7
US-09-925-300-330
; Sequence 330, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 330
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (643)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (657)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (685)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-330

Alignment Scores:
Pred. No.: 2.55e-36 Length: 696
Score: 352.00 Matches: 67
Percent Similarity: 98.53% Conservative: 0
Best Local Similarity: 98.53% Mismatches: 1
Query Match: 98.05% Indels: 0
DB: 10 Gaps: 0

US-09-830-144-4_COPY_437_504 (1-68) x US-09-925-300-330 (1-696)
QY 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSerSer 20
Db 49 CAAAGCCCGACCTTAACCTGACGTCACCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 108
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGluAspGly 40
Db 109 TCTGACGAGGCGCTCTTCCGCTCCGCGCCGCCCACTCGCTCCCGCTGCGAGGAGCGGT 168
QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTyrPseValAspHisGly 60
Db 169 CGTGTGAGCCCTATGATGAGACTTGTGAGATTTCACGCCCTCTGAGACGTCGACATGCG 228
QY 61 GluGlnSerValValThrAlaPro 68
Db 229 GAGCAGAGCGGTGTGACAGCACCG 252

RESULT 8
US-09-974-300-6219
; Sequence 6219, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6219
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-6219

Alignment Scores:
Pred. No.: 2.44 Length: 996
Score: 71.00 Matches: 19
Percent Similarity: 50.00% Conservative: 13
Best Local Similarity: 29.69% Mismatches: 24
Query Match: 19.78% Indels: 8
DB: 10 Gaps: 3
```

```
;
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 17981/111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1560 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..1541
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 30..1541
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-123-427-5
```

```
Alignment Scores:
Pred. No.: 9.07e-37 Length: 1560
Score: 359.00 Matches: 68
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0
```

US-09-830-144-4\_COPY\_437\_504 (1-68) x US-10-123-427-5 (1-1560)

```
Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20
Db 1338 CAAAGCCCGACCTTAACCCCTGCAGTCCACCAACGACGACGAGGAGCGTCCAGC 1397
Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40
Db 1398 TCTGACGAGCGCTCTTCCGCTCCCGCCGCCCACTCGCTCCCGCTGGCGAGGCGGT 1457
Qy 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60
Db 1458 CGTGTGGAGCCCTATGTGGACTTTGCTGAGTTTACCGCCCTCGGAGCGTGACCATGCG 1517
Qy 61 GluGlnSerValValThrAlaPro 68
Db 1518 GAGCAGAGCGGTGTGACAGCACCG 1541
```

```
RESULT 4
US-10-158-895-42
; Sequence 42, Application US/10158895
; Patent No. US20020155624A1
; GENERAL INFORMATION:
; APPLICANT: ONO, KOICHIRO
; APPLICANT: TSUCHIYA, MASAYUKI
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES
; FILE REFERENCE: 053466/0278
; CURRENT APPLICATION NUMBER: US/10/158,895
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/529,279
; PRIOR FILING DATE: 2000-04-11
; PRIOR FILING DATE: 1998-10-22
; PRIOR FILING DATE: 1997-10-22
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 42
; LENGTH: 1568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
```

```
; LOCATION: (11)..(1549)
US-10-158-895-42
Alignment Scores:
Pred. No.: 9.12e-37 Length: 1568
Score: 359.00 Matches: 68
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-830-144-4_COPY_437_504 (1-68) x US-10-158-895-42 (1-1568)
Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20
Db 1346 CAAAGCCCGACCTTAACCCCTGCAGTCCACCAACGACGACGAGGAGCGTCCAGC 1405
Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40
Db 1406 TCTGACGAGCGCTCTTCCGCTCCCGCCGCCCACTCGCTCCCGCTGGCGAGGCGGT 1465
Qy 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTrpSerValAspHisGly 60
Db 1466 CGTGTGGAGCCCTATGTGGACTTTGCTGAGTTTACCGCCCTCGGAGCGTGACCATGCG 1525
Qy 61 GluGlnSerValValThrAlaPro 68
Db 1526 GAGCAGAGCGGTGTGACAGCACCG 1549
RESULT 5
US-10-158-895-10
; Sequence 10, Application US/10158895
; Patent No. US20020155624A1
; GENERAL INFORMATION:
; APPLICANT: ONO, KOICHIRO
; APPLICANT: OHYOMO, TOSHIHIKO
; APPLICANT: TSUCHIYA, MASAYUKI
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES
; FILE REFERENCE: 053466/0278
; CURRENT APPLICATION NUMBER: US/10/158,895
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/529,279
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: PCT/JP98/04796
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: JP 9/290188
; PRIOR FILING DATE: 1997-10-22
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 10
; LENGTH: 1569
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1557)
US-10-158-895-10
Alignment Scores:
Pred. No.: 9.13e-37 Length: 1569
Score: 359.00 Matches: 68
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-830-144-4_COPY_437_504 (1-68) x US-10-158-895-10 (1-1569)
Qy 1 GlnSerProThrLeuThrLeuGlnSerThrAsnThrHisThrGlnSerSerSerSer 20
Db 1315 CAAAGCCCGACCTTAACCCCTGCAGTCCACCAACGACGACGAGGAGCGTCCAGC 1374
Qy 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProGlyGluAspGly 40
```

Alignment Scores:  
Pred. No.: 9.07e-37 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x US-10-158-895-1 (1-1560)

QY 1 GlnSerProThleuThrluInserThraAnthHisThrGlnSerSerSerSer 20  
Db 1338 CAAGGCCGACCTTAACCTTGACATCCCAACGACGACGACGACGACCTCCAGC 1397  
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGlyAspGly 40  
Db 1398 TCTGACGAGGAGGCTCTTCCGCTCCGCGCCGACCACTCGCTCCGCGCTGCGAGGACGGT 1457

QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTyrSerValAspHisGly 60  
Db 1458 CGGTGGAGCCCTTAAGTGGACCTTCTGAGTTTACCGCCTCTGAGACCTGGACCATGGC 1517

QY 61 GluGlnSerValValThrAlaPro 68  
Db 1518 GAGCAGACCGGTGGTGACAGCACCG 1541

RESULT 2  
US-10-123-427-1  
Sequence 1, Application US/10123427  
Patent No. US20020119525A1  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihiko  
NISHIDA, Eisuke  
TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/123,427  
FILING DATE: 17-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/406,854  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US/08/752,891  
FILING DATE: 20-NOV-1996  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 17981/111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1560 base pairs  
TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1541  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 30..1541  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-123-427-1

Alignment Scores:  
Pred. No.: 9.07e-37 Length: 1560  
Score: 359.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-09-830-144-4\_COPY\_437\_504 (1-68) x US-10-123-427-1 (1-1560)

QY 1 GlnSerProThleuThrluInserThraAnthHisThrGlnSerSerSerSer 20  
Db 1338 CAAGGCCGACCTTAACCTTGACATCCCAACGACGACGACGACGACCTCCAGC 1397  
QY 21 SerAspGlyGlyLeuPheArgSerArgProAlaHisSerLeuProProGlyGlyAspGly 40  
Db 1398 TCTGACGAGGAGGCTCTTCCGCTCCGCGCCGACCACTCGCTCCGCGCTGCGAGGACGGT 1457

QY 41 ArgValGluProTyrValAspPheAlaGluPheTyrArgLeuTyrSerValAspHisGly 60  
Db 1458 CGGTGGAGCCCTTAAGTGGACCTTCTGAGTTTACCGCCTCTGAGACCTGGACCATGGC 1517

QY 61 GluGlnSerValValThrAlaPro 68  
Db 1518 GAGCAGACCGGTGGTGACAGCACCG 1541

RESULT 3  
US-10-123-427-5  
Sequence 5, Application US/10123427  
Patent No. US20020119525A1  
GENERAL INFORMATION:  
APPLICANT: MATSUMOTO, Kunihiko  
NISHIDA, Eisuke  
TITLE OF INVENTION: TAB1 PROTEIN AND DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/123,427  
FILING DATE: 17-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/406,854  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US/08/752,891  
FILING DATE: 20-NOV-1996  
APPLICATION NUMBER: JP 8-300856  
FILING DATE: 28-OCT-1996  
APPLICATION NUMBER: JP 8-126282  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 10, 2002, 05:51:41 ; Search time 86 Seconds

(without alignments)

308.319 Million cell updates/sec

Title: US-09-830-144-4\_COPY\_437\_504

Perfect score: 359

Sequence: 1 OSPTILTOSTNTHQTSSSSS.....AEFYRLWSVDHGQSVWTP 68

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Zgapop 6.0 , Zgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 350425 seqs, 194966369 residues

Total number of hits satisfying chosen parameters: 700850

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DEV=xlp

-Q=/cgn2\_1/USPTO\_spool/US09830144/runat\_04122002\_141844\_5032/app\_query.fasta.1.263

-DB=PublishedApplications NA -Q=FASTAP -SUFFIX=p2n.rnpb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62

-TRANS-human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORES=ptc -THR MAX=100

-THR\_MIN=0 -ALIGN=15 -MOD=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0

-MAXLEN=200000000 -USER=US09830144 @CGN 1 1 36 @runat\_04122002\_141844\_5032

-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG

-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6

-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7-

Database :

Published Applications NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*

2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*

3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*

4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*

5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*

6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*

7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*

8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*

9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*

10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*

11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*

12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*

13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*

14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	359	100.0	1560	9	US-10-158-895-1
2	359	100.0	1560	12	US-10-123-427-1
3	359	100.0	1560	12	US-10-123-427-5
4	359	100.0	1568	9	US-10-158-895-42

5	359	100.0	1569	9	US-10-158-895-10
6	359	100.0	16877	10	US-09-764-877-3349
7	352	98.1	696	10	US-09-925-300-330
8	71	19.8	996	10	US-09-974-300-6219
C 9	69.5	19.4	1492	9	US-09-941-831-7
C 10	68	18.9	273	10	US-09-864-761-22569
C 11	68	18.9	8918	10	US-09-764-864-1682
C 12	68	18.9	8919	10	US-09-764-864-1683
C 13	64.5	18.0	966	9	US-09-338-842A-2555
14	64.5	18.0	42432	9	US-10-027-806-2
15	63.5	17.7	1101	9	US-09-852-797-23
16	63.5	17.7	1101	10	US-09-853-161-23
17	63.5	17.7	1101	10	US-09-852-659A-23
18	63.5	17.7	1473	9	US-09-852-797-43
19	63.5	17.7	1473	10	US-09-853-161-43
20	63.5	17.7	1473	10	US-09-852-659A-43
C 21	63.5	17.7	2784	10	US-09-205-658-39
C 22	63.5	17.7	2784	10	US-09-844-353A-39
C 23	63.5	17.7	3017	10	US-09-205-658-52
C 24	63.5	17.7	3017	10	US-09-844-353A-52
C 25	63.5	17.7	3119	10	US-09-205-658-53
C 26	63.5	17.7	3119	10	US-09-844-353A-53
27	63.5	17.7	3236	10	US-09-954-456-43
C 28	63.5	17.7	42450	10	US-09-815-048-3
29	62.5	17.4	167343	10	US-09-962-436-281
30	62.5	17.4	167343	10	US-09-964-824A-273
C 31	62	17.3	447	10	US-09-864-761-10872
C 32	62	17.3	1161	10	US-09-371-307-77
C 33	62	17.3	2637	10	US-09-815-242-9859
C 34	61.5	17.1	245	10	US-09-960-352-12511
35	61	17.0	1149	10	US-09-764-847-330
36	61	17.0	1450	10	US-09-764-853-186
37	61	17.0	3320	10	US-09-862-658-1
38	61	17.0	4536	10	US-09-801-368-249
C 39	60.5	16.9	172	10	US-09-864-761-27512
C 40	60	16.7	429	10	US-09-960-352-14704
41	60	16.7	509	9	US-09-764-868-255
42	60	16.7	510	9	US-09-764-868-566
C 43	60	16.7	584	10	US-09-864-761-15289
C 44	60	16.7	900	10	US-09-982-809-21
45	60	16.7	1606	10	US-09-853-386-55

#### ALIGNMENTS

RESULT 1  
US-10-158-895-1  
; Sequence 1, Application US/10158895  
; Patent No. US20020155624A1  
; GENERAL INFORMATION:  
; APPLICANT: ONO KOICHIRO  
; APPLICANT: OHTOMO, TOSHIHIKO  
; APPLICANT: TSUCHIYA, MASAYUKI  
; TITLE OF INVENTION: METHOD OF SCREENING TGF-BETA INHIBITORY SUBSTANCES  
; FILE REFERENCE: 053466/0278  
; CURRENT APPLICATION NUMBER: US/10/158,895  
; CURRENT FILING DATE: 2002-06-03  
; PRIOR APPLICATION NUMBER: US/09/529,279  
; PRIOR FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: PCT/JP98/04796  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: JP 9/290188  
; PRIOR FILING DATE: 1997-10-22  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1560  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (30)...(1541)  
US-10-158-895-1